Indiana Law Review

Volume 22 No. 3 1989

ARTICLES

The Efficiency of a Disgorgement as a Remedy for Breach of Contract
Sidney W. DeLong

Defending Purchase Money Security Interests Under Article 9 of the UCC From Professor Buckley
Paul M. Shupack

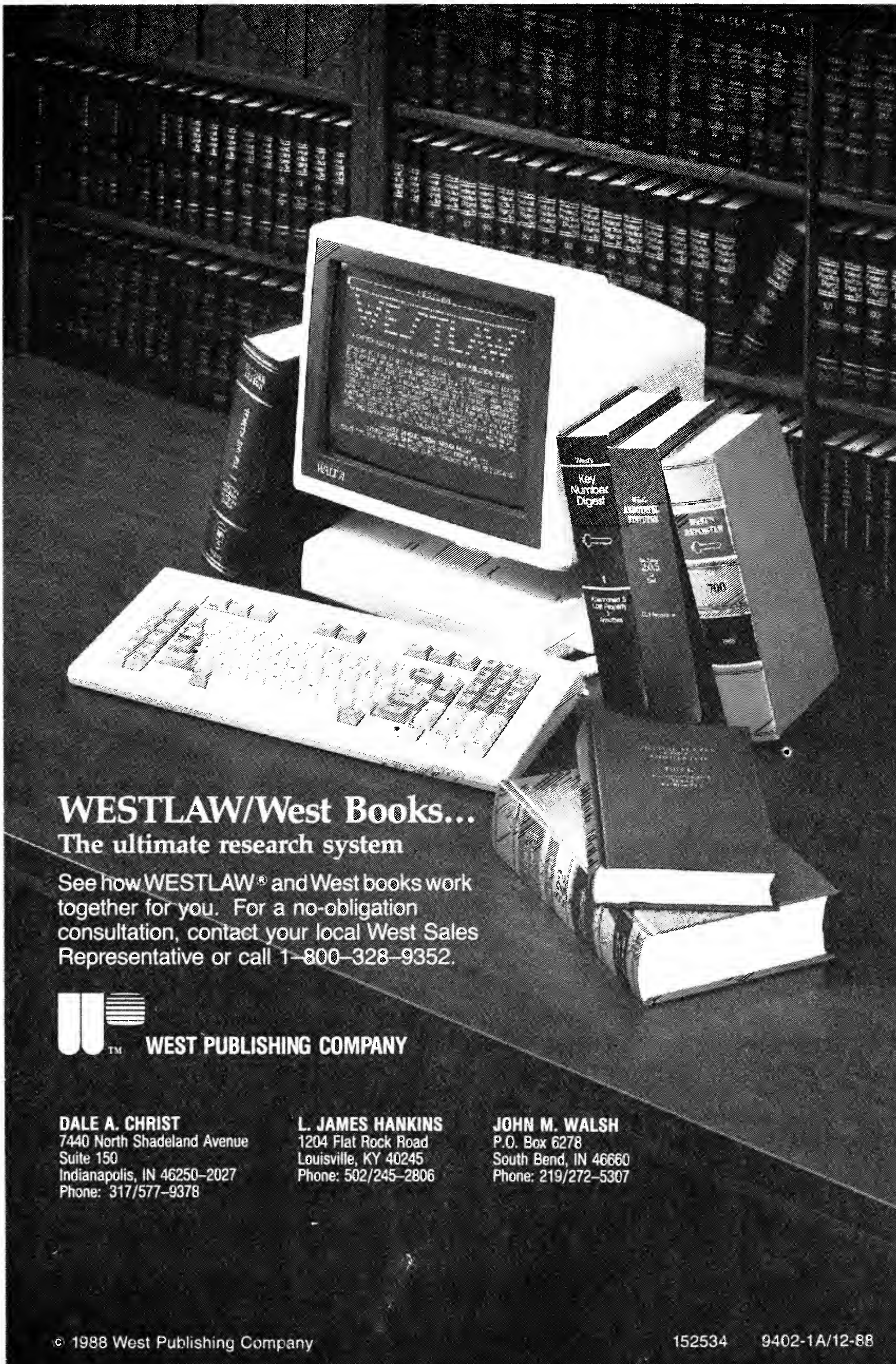
NOTES

Random Drug Testing of Police Officers: A Proposed Procedure Which Satisfies Fourth Amendment Requirements

Delimiting the Manufacturer's Liability: An Examination of Loss of Consortium Recovery in Strict Products Liability Actions Under Section 402A of the Restatement (Second) of Torts

The Test for Patent Infringement Under the Doctrine of Equivalents After *Pennwalt v. Durand-Wayland*

The Standard of Proof in Civil RICO Actions for Treble Damages: Why the Clear and Convincing Standard Should Apply


Key Number Digest
WESTLAW/West Books...
The ultimate research system
See how WESTLAW® and West books work together for you. For a no-obligation consultation, contact your local West Sales Representative or call 1–800–328–9352.
WEST PUBLISHING COMPANY
DALE A. CHRIST
7440 North Shadeland Avenue
Suite 150
Indianapolis, IN 46250–2027
Phone: 317/577–9378
L. JAMES HANKINS
1204 Flat Rock Road
Louisville, KY 40245
Phone: 502/245–2806
JOHN M. WALSH
P.O. Box 6278
South Bend, IN 46660
Phone: 219/272–5307
© 1988 West Publishing Company
152534 9402-1A/12-88

Indiana National

on the subject of options

Investment Option	Minimum Investment	Maturities	Credit	Income	Benefits
Tax-Exempt Municipal Bonds/ Unit Investment Trusts	$ 5,000 1,000	To 30 Years	Backed by Issuer	Semi-Annual Coupon/ Monthly, Quarterly, Semi-Annual	Exempt from Federal Income Tax for Corporations, and from State and Federal Tax for Individuals.*
Money Market Funds	1,000	Day to Day	Issuing Fund	Monthly	High current income, preservation of capital, easy withdrawal procedures. Share value stability & check writing redemption.
Mutual Funds	250	Open End Funds	Issuing Fund	Monthly	High current return, monthly income, liquidity, high quality portfolios.
U.S. Treasury Bills	10,000	One Year or less	Direct Obligations of U.S. Treasury	Discount to Maturity	High quality, extremely liquid, state tax exempt.*
U.S. Treasury Notes	5,000 1,000	1 - 3 Years 3 - 10 Years	Direct Obligations of U.S. Treasury	Semi-Annual Coupon	High quality, extremely marketable, state tax exempt.*
U.S. Treasury Bonds	1,000	Over 10 Years	Direct Obligations of U.S. Treasury	Semi-Annual Coupon	High quality, extremely marketable, state tax exempt.*
Discount Notes: Primarily Federal Home Loan Banks, FNMA, Federal Farm Credit Banks	10,000	30 to 360 Days	Issuing Agency	At Maturity	May offer higher yields and smaller minimums than Treasury issues. High quality and extremely liquid.*
Federal Farm Credit, FNMA, Federal Home Loan Banks	1,000 5,000 10,000	6 Months to 30 Years	Issuing Agency	At Maturity or Semi-Annual Coupon	May offer higher yields and smaller minimums than Treasury issues. High quality and good marketability.*
Bankers Acceptances	100,000	10 Days to 270 Days	Guaranteed by Issuing Bank	Discount to Maturity	High quality and extremely liquid.*
Commercial Paper	100,000	To 270 Days	Backed by Issuer	At Maturity or Discount to Maturity	High yields, good marketability, highly liquid.*

*Possible market risk if sold before maturity date.

Have you considered all your options?

Investment banking from Indiana National

You may not have thought of Indiana National as a source of investment banking. But we're one of Indiana's major marketers of tax-exempt municipal bonds and notes, government and agency securities.

Reaching your investment goals

Our account executives have proven their capabilities and market insights by assisting clients in the acquisition of over $32 billion in fixed income securities during the last five years.

And because they focus on safety, liquidity and rate of return, they provide the kind of securities that meet clients' investment objectives.

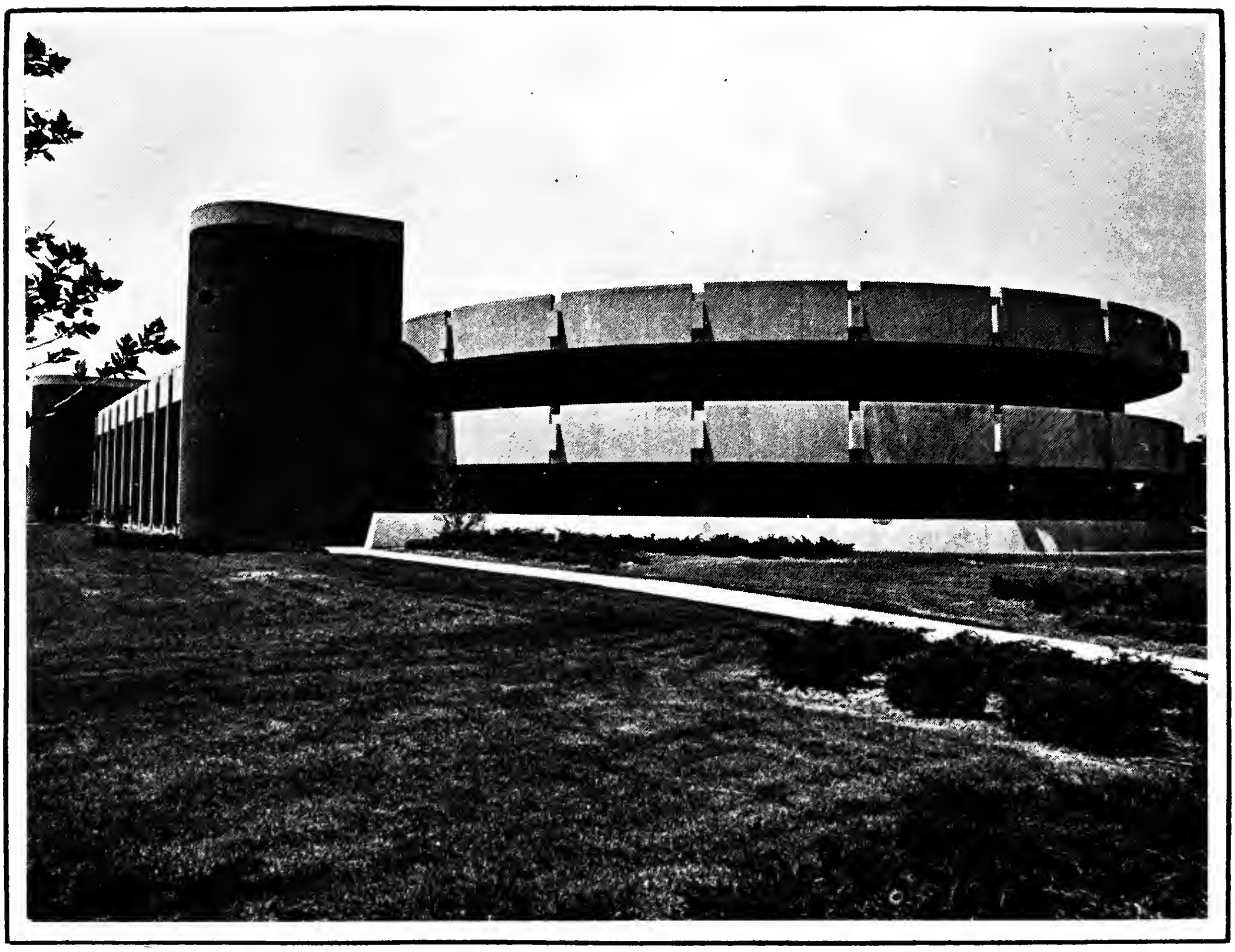
Managing your investment plan

At Indiana National you receive individualized attention to your investment goals. And you benefit from our decades of experience in personal service and prompt execution of trades.

The next time you're ready to make an investment, think of Indiana National. Because, when it comes to meeting your investment objectives, we've got all the options.

Call 1-800-382-1102 or (317) 266-6829 for more information.

Indiana National.
Investment Banking Division Member FDIC



Please enter my subscription to the
INDIANA LAW REVIEW

NAME ______________________________

ADDRESS ______________________________

____

____

Enclosed is $______ for ____ subscriptions.

Bill me for ______ subscriptions.

Mail to:

INDIANA LAW REVIEW
INDIANA UNIVERSITY
SCHOOL OF LAW—INDIANAPOLIS
735 West New York Street
Indianapolis, Indiana 46202

Subscription Rates (one year):

Regular, $22; Foreign, $25; Survey, $15;
Single Issue, $7

Indiana Law Review

Volume 22 | 1989 | Number 3

TABLE OF CONTENTS

Articles

The Efficiency of a Disgorgement as a Remedy for Breach of Contract .... *Sidney W. DeLong* 737

Defending Purchase Money Security Interests Under Article 9 of the UCC From Professor Buckley .... *Paul M. Shupack* 777

NOTES

Random Drug Testing of Police Officers: A Proposed Procedure Which Satisfies Fourth Amendment Requirements 799

Delimiting the Manufacturer's Liability: An Examination of Loss of Consortium Recovery in Strict Products Liability Actions Under Section 402A of the Restatement (Second) of Torts 821

The Test for Patent Infringement Under the Doctrine of Equivalents After *Pennwalt v. Durand-Wayland* 849

The Standard of Proof in Civil RICO Actions for Treble Damages: Why the Clear and Convincing Standard Should Apply 881

Volume 22 | 1989 | Number 3

The INDIANA LAW REVIEW (ISSN 0090-4198) is the property of Indiana University and is published quarterly by the Indiana University School of Law—Indianapolis, which assumes complete editorial responsibility thereof. Subscription rates: one year $22; foreign $25. Please notify us one month in advance of any change in address and include both old and new addresses with zip codes to ensure delivery of all issues. Send all correspondence to Editorial Assistant, Indiana Law Review, Indiana University School of Law—Indianapolis, 735 West New York Street, Indianapolis, Indiana 46202. Publication office: 735 West New York Street, Indianapolis, Indiana 46202. Second class postage paid at Indianapolis, Indiana 46201.

POSTMASTER: Send address changes to INDIANA LAW REVIEW, 735 West New York Street, Indianapolis, Indiana 46202.

Indiana Law Review

Volume 22 1989

Indiana University School of Law—Indianapolis
1988-89 ADMINISTRATIVE OFFICERS AND FACULTY

Administrative Officers

Thomas M. Ehrlich, *LL.B., President of the University*
Gerald L. Bepko, *LL.M., Vice-President*
Norman Lefstein, *LL.B., Dean*
James F. Bindley, *J.D., Assistant Dean for Administration*
Debra A. Falender, *J.D., Acting Assistant Dean of Administration*
Lawrence P. Wilkins, *LL.M., Associate Dean for Academic Affairs*

Faculty

Thomas B. Allington, *Professor. B.S., University of Nebraska, 1964; J.D., 1966; LL.M., New York University, 1971.*

Edward P. Archer, *Professor. B.M.E., Rensselaer Polytechnic Institute, 1958; J.D., Gerogetown University, 1962; LL.M., 1964.*

James F. Bailey, III., *Professor and Director of Law Library. A.B., University of Michigan, 1961; J.D., 1964; M.A.L.S., 1970.*

Gerald L. Bepko, *Vice President and Professor. B.S., Northern Illinois University, 1962; J.D., IIT/Chicago-Kent College of Law, 1965; LL.M., Yale University, 1972.*

James F. Bindley, *Assistant Dean for Administration and Director of Placement & Development, B.A., Loyola University, 1969; J.D., University of Kentucky. 1972.*

Paul N. Cox, *Professor. B.S., Utah State University, 1972; J.D., University of Utah, 1974; LL.M., University of Virginia, 1980.*

Clyde Harrison Crockett, *Professor. A.B., University of Texas, 1962; J.D., 1965; LL.M., University of London (The London School of Economics and Political Science), 1972.*

Debra A. Falender, *Professor. A.B., Mount Holyoke College, 1970; J.D., Indiana University, 1975.*

David A. Funk, *Professor. A.B., College of Wooster, 1949; J.D., Case Western Reserve University, 1951; M.A., The Ohio State University 1968; LL.M., Case Western Reserve University, 1972; LL.M., Columbia University, 1973.*

Paul J. Galanti, *Professor. A.B., Bowdoin College, 1960; J.D., University of Chicago, 1963.*

Helen P. Garfield, *Professor. B.S.J., Northwestern University, 1945; J.D., University of Colorado, 1967.*

Harold Greenberg, *Associate Professor. A.B., Temple University, 1959; J.D., University of Pennsylvania, 1962.*

Jeffrey W. Grove, *Professor. A.B., Juniata College, 1965; J.D., George Washington University, 1969.*

William F. Harvey, *Carl M. Gray Professor of Law. A.B., University of Missouri, 1954; J.D., Georgetown University, 1959; LL.M., 1961.*

Paul T. Hayden, *Assistant Professor. B.A., Yale University, 1979; J.D., University of California, Los Angeles, 1984.*

W. William Hodes, *Professor. A.B., Harvard College, 1966; J.D., Rugters Newark, 1969.*

Lawrence A. Jegen, III., *Thomas F. Sheehan Professor of Tax Law and Policy, 1982. A.B., Beloit College 1956; J.D., The University of Michigan 1959; M.B.A., 1960, LL.M., Harvard University, 1963.*

Henry C. Karlson, *Professor. A.B., University of Illinois, 1965; J.D., 1968; LL.M., 1977.*

William Andrew Kerr, *Professor. A.B., West Virginia University, 1955. J.D., 1957, LL.M., Harvard University, 1958; D.B., Duke University, 1968.*

Eleanor D. Kinney, *Assistant Professor. A.B., Duke Universty, 1969; M.A., University of Chicago, 1970; J.D., Duke University, 1973.*

Walter W. Krieger, *Associate Professor. A.B., Bellarmine College, 1959; J.D., University of Louisville, 1962; LL.M., George Washington University, 1969.*

Norman Lefstein, *Dean and Professor. LL.B., University of Illinois, 1961; LL.M., Georgetown University, 1964.*

David P. Leonard, *Professor. B.A., University of California at San Diego, 1974; J.D., UCLA School of Law, 1977.*

William E. Marsh, *Professor. B.S., University of Nebraska, 1965; J.D., 1958.*

SUSANAH M. MEAD, *Associate Professor. B.A., Smith College, 1969; J.D., Indiana University, 1976.*

MARY H. MITCHELL, *Associate Professor. A.B., Butler University, 1975; J.D., Cornell Law School, 1978.*

DAVID R. PAPKE, *Associate Professor. A.B., Harvard College, 1969; J.D., Yale Law School, 1973; M.A., in American Studies, Yale University, 1973; Ph.D., in American Studies, The University of Michigan, 1984.*

RONALD W. POLSTON, *Professor. B.S., Eastern Illinois University, 1953; LL.B., University of Illinois, 1958.*

KENNETH M. STROUD, *Professor. A.B., Indiana University, 1958; J.D., 1961.*

JAMES W. TORKE, *Professor. B.S., University of Wisconsin, 1963; J.D., 1968.*

JOE A. TUCKER, *Assistant Professor. B.A., Houston, 1977; J.D., University of Texas, 1981.*

JAMES PATRICK WHITE, *Professor of Law. A.B., University of Iowa, 1953; J.D., 1956; LL.M., George Washington University, 1959.*

LAWRENCE P. WILKINS, *Professor. B.A., The Ohio State University, 1968; J.D., Capital University Law School, 1973; LL.M., University of Texas School of Law, 1974.*

MARY WOLF, *Assistant Professor. B.A., Saint Xavier College, 1969; J.D., University of Iowa College of Law, 1974.*

Emeriti

AGNES P. BARRETT, *Associate Professor Emeritus. B.S., Indiana University, 1942; J.D., 1964.*

CLEON H. FOUST, *Professor Emeritus. A.B., Wabash College, 1928; J.D., University of Arizona, 1933.*

JOHN S. GRIMES, *Professor of Jurisprudence Emeritus. A.B., Indiana University. 1929; J.D., 1931.*

MELVIN C. POLAND, *Cleon H. Foust Professor of Law Emeritus. B.S., Kansas State University, 1940; LL.B., Washburn University, 1949; LL.M., The University of Michigan, 1950.*

R. BRUCE TOWNSEND, *Cleon H. Foust Professor of Law Emeritus, A.B., Coe College, 1938; J.D., University of Iowa, 1940.*

Legal Writing Instructors

CYNTHIA ADAMS, *B.A., Indiana State University, 1976; B.A., Kentucky Wesleyan College, 1977; J.D., Indiana University, 1983.*

ROBERT L. ELLIS, *B.A., Ohio State University, 1971; J.D., Antioch School of Law, 1981.*

VICKIE RENFROW, *Lecturer, B.A., University of Northern Iowa, 1970; M.A., 1971; Ph.D., Indiana University, Bloomington, 1976; J.D., Indiana University, Bloomington, 1981.*

JOAN RUHTENBERG, *Director of Legal Writing. B.A., Mississippi University Women, 1959; J.D., Indiana University, 1980.*

Law Library Staff

MARY P. HUDSON, *Assistant Librarian, B.A., Ball State, 1979; M.L.S., Indiana University, 1973.*

WENDELL F. JOHNTING, *Technical Services Librarian, A.B., Taylor University, 1974; M.L.S., Indiana University, 1975.*

KIYOSKI OTSU, *Assistant Librarian, Parkland College, A.A., 1976; A.B., University of Illinois, 1980; M.S., 1982; C.A.S., 1983.*

TERRI L. ROSS, *Assistant Librarian, B.A., Indiana University, 1982; M.L.S., 1983.*

MERLIN P. WHITEMAN, *Readers' Services Librarian, B.A., Hope College, 1973; M.L.S., Indiana University, 1974; J.D., 1982.*

Indiana Law Review

Volume 22 | 1989 | Number 3

THE EFFICIENCY OF A DISGORGEMENT AS A REMEDY FOR BREACH OF CONTRACT

Sidney W. DeLong*

Thesis: Economic analysis suggests that to give a contract promisee a general remedy that would require a breaching promisor to disgorge any benefit of breach would hinder the efficient post-contractual reallocation of performance resources. This article explores certain situations in which disgorgement appears to be an efficient remedy for breach of contract, including cases in which the breaching party refuses to pay contract damages at the time of breach. A rule permitting promisees to recover as "prejudgment interest" the breacher's benefit from withholding payment of damages would, in theory, be efficient in allocating the risk of the breacher's credit worthiness to the best risk bearer. Colorado has implemented this remedy, but has, over time, extended it to require the defendant to disgorge the benefits of breach instead of the benefits of withholding damages. This extension appears to be unwarranted. Moreover, serious practical and theoretical difficulties in defining and measuring the benefits of either breaching or withholding damages cast doubt on the utility of the disgorgement remedy.

CONTENTS

I. INTRODUCTION 738
II. REMEDY EFFICIENCY AS COST AVOIDANCE 740
  A. *The Cost of Contract Remedies* 740
  B. *Efficient Breach of Contract* 742
  C. *The Efficiency of Traditional Disgorgement Remedies for Breach of Contract* 746

* Assistant Professor of Law, University of Puget Sound School of Law, B.A. Vanderbilt University 1969, J.D., Yale, 1974.

D. *The Efficiency of Disgorgement as a Remedy for Conversion* 751
E. *Disgorgement Reconsidered* 752
III. DAMAGES FOR FAILURE TO PAY DAMAGES 755
A. *Ex Ante Cost Avoidance Analysis of Prejudgment Interest* 756
B. *Post-Breach Cost Avoidance Analysis of Prejudgment Interest* 757
1. *Compensation Method* 757
2. *Disgorgement Method* 758
C. *The Effect of Borrowing Rates* 759
IV. A CASE STUDY IN DAMAGES AS INTEREST 760
V. THE PROBLEM OF CAUSATION: LEGAL CONCEPTUALISM AND ECONOMIC ANALYSIS 768
A. *Calculation of the Benefits From Reallocating Performance Resources* 769
1. *Identified Performance Resources* 769
2. *Non-Identified Performance Resources* 771
3. *Measuring Benefit by "Savings" Resulting from Breach* 771
4. *Informational Transaction Costs of Calculating the Breacher's Benefit* 772
B. *Calculation of the Benefits from Withholding Payment of Damages* 773
VI. CONCLUSION 775

I. INTRODUCTION

It is by now a commonplace that the law of contract remedies should take efficiency[1] into account by encouraging contract parties to

1. A definition of efficiency often used in legal analysis is that used by Posner in the second edition of his treatise: "'Efficiency' means exploiting economic resources in such a way that, 'value'—human satisfaction *as measured by aggregate consumer willingness to pay* for goods and services—is maximized." R. POSNER, ECONOMIC ANALYSIS OF LAW 10 (2d ed. 1977) (Emphasis in original). Posner employs the Kaldor-Hicks concept of efficiency with respect to voluntary transactions: a transaction is efficient if it leads to a reallocation of resources that increases net social value as measured by willingness to pay. *Id.*; Coleman, *Efficiency, Utility, and Wealth Maximization*, 8 HOFSTRA L. REV. 509, 513-14 (1980). *See also* R. POSNER, ECONOMIC ANALYSIS OF LAW 106-08 (3d ed. 1986).

This article will differentiate between transactional efficiency, or the net gain in wealth from a single exchange, and rule efficiency, or the tendency of a rule to establish incentives that would lead parties to act in efficient ways. Inefficient transactions can occur under efficient rules.

allocate resources to their highest valued uses even after the contract is formed. Whether or not efficiency should be the only criterion of contract remedies,[2] it remains at least a relevant topic of inquiry.

The usual remedy for breach of contract, expectation damages, is designed, with some notable limitations,[3] to put the non-breaching party (the promisee) in the position he would have enjoyed if the contract had been performed,[4] and to permit the breaching party (the promisor) to retain any excess benefit obtained as a result of the breach.[5] This article explores some of the efficiency ramifications of a remedy that would require a breaching party who refuses to pay damages at the time of breach to disgorge the benefits of the breach to the non-breaching party. It also examines some of the conceptual problems in defining the "benefits of breach."

Part I analyzes the efficiency of disgorgement as both a contract and tort remedy. Cost avoidance analysis suggests that disgorgement is inefficient as a general remedy for breach of contract because it inhibits efficient reallocation of performance resources by requiring

---

2. For a sample of the controversy over the role of efficiency as a normative value, *see* Posner, *The Ethical and Political Basis of the Efficiency Norm in Common Law Adjudication*, 8 Hofstra L. Rev. 487 (1980); Coleman, *supra* note 1, at 531-51; Dworkin, *Why Efficiency? A Response to Professors Calabresi and Posner*, 8 Hofstra L. Rev. 563 (1980); Kornhauser, *A Guide to the Perplexed Claims of Efficiency in the Law*, 8 Hofstra L. Rev. 591 (1980).

3. *See infra* notes 47-56 and accompanying text.

4. Restatement (Second) of the Law of Contracts § 344(a) (1979) [hereinafter cited as Restatement (Second) of Contracts]. *See also* U.C.C. § 1-106(1) (1977) ("The remedies provided by this Act shall be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed, but neither consequential or special nor penal damages may be had except as specifically provided in this Act or by other rule of law."); 5 A. Corbin, Corbin on Contracts §§ 992, 1002 (1964) ("The effort is made to put the injured party in as good a position as he would have been put by full performance of the contract, at the least cost to the defendant and without charging him with harms that he had no sufficient reason to foresee when he made the contract."); *Id.* at § 1002.; S. Williston & G. Thompson, Contracts § 1338 (rev. ed. 1937).

The promisee's interest in performance (his expectation interest), is contrasted to alternative bases of recovery: restitution (the return of benefits conferred upon a breaching party) and reliance (compensation for amounts spent in reliance upon the promised performance). *See* Fuller & Perdue, *The Reliance Interest in Contract Damages: 1* 46 Yale L.J. 52 (1936). Briefly described, reliance damages and restitution damages put the promisee and promisor, respectively, in the positions they would have enjoyed had the contract not been entered. Farnsworth, *Legal Remedies For Breach of Contract*, 70 Colum. L. Rev. 1145, 1148 (1970).

In addition to money damages, contract parties are sometimes entitled to specific performance of the contract promise. *See* Kronman, *Specific Performance*, 45 U. Chi. L. Rev. 351 (1978).

5. R. Posner, Economic Analysis of Law 107-08 (3d ed. 1986).

contracting parties to negotiate before efficient breach can take place. By contrast, disgorgement appears to be an efficient remedy for the tort of conversion, where the possible benefit of the convertor's efficient reallocation of converted goods is outweighed by the benefit of the certainty gained when the convertor is required to establish the efficiency of the reallocation in a voluntary transaction. This rationale, premised on the uncertainty of judicial determinations of value, might also justify a disgorgement remedy in some contract breach situations. Two such situtations are analyzed.

Part II turns to the question of whether a contract breacher may "efficiently" withhold payment of contract damages. It compares the efficiency of using expectation and disgorgement measures for calculating prejudgment interest. Here, too, cost avoidance analysis suggests that a disgorgement remedy is more rule efficient than an expectation damages remedy in calculating prejudgment interest.

Part III describes the statutory and case law of a jurisdiction that has adopted a disgorgement method of calculating prejudgment interest. It also traces the history whereby that remedy has been enlarged so as to permit a plaintiff to recover the benefits of efficient breach instead of the benefits of withholding damages. The results of this evaluation are troublesome and suggest a practical inability to contain a prejudgment interest disgorgement remedy to its theoretical conceptual limits.

Part IV seeks the source of this inability by analyzing the assumption that the "benefit" caused by either a breach of contract or a withholding of damages can be determined by traditional legal analysis. Inherent limitations on the causal reasoning employed by courts will increase the uncertainty and cost of the disgorgement remedy, impairing its ability to achieve its theoretical efficiency.

II. REMEDY EFFICIENCY AS COST AVOIDANCE

A. *The Cost of Contract Remedies*

Like the rest of the substantive law of contracts, the law of contract remedies can be seen as a set of ready-made contract terms supplied by the state to govern the contracts of parties who do not otherwise agree to specific remedy terms.[6] Because the law of contract remedies will directly affect the risks and rewards of contracting, it is reasonable

---

6. Goetz & Scott, *The Mitigation Principle: Toward a General Theory of Contractual Obligation*, 69 VA. L. REV. 967, 971 (1983). *See also* Posner & Rosenfield, *Impossibility and Related Doctrines in Contract Law: An Economic Analysis*, 6 J. LEGAL STUD. 83, 88-92 (1977).

to assume that rational parties will take those remedies into account when they negotiate the price of contractual promises.[7] A promisee will pay more for additional contract remedies,[8] while a promisor will demand a contract price that reflects the full costs of the risks that the law allocates to her[9] in its remedy provisions.[10]

When, for whatever reason,[11] the parties have failed to agree about the remedies for a breach that comes to pass, economic analysts have argued that the law should impose the remedy that the parties would have agreed upon.[12] In the absence of evidence about actual agreement, it is assumed that the parties would have selected remedies that would minimize the joint costs of the bargain.[13]

The cost analysis method of recreating the remedy the parties would have selected posits that a promisor would be willing to sell any remedy

---

7. *See* Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 S. Cal. L. Rev. 629, 642 (1988). Critical to this step are the assumptions that parties are perfectly informed about the law and the costs that contract remedies will impose, and that they can bargain over the forms and contract prices of the remedies costlessly.

8. The value of remedies is only one component of the value of a contractual promise. The value of the contract to the promisee is the combination of the value of the promisee's expectation that the promisor will perform as promised and the expected value of the remedy available upon breach. These two elements are interrelated in complex ways. Thus, the relative probabilities of performance and breach will depend in part on the severity of the remedy the promisor expects to pay in the event of breach. The value of performance will often depend upon the degree to which the promisee can rely on the promised performance, which will in turn depend upon the expected value of the remedy in the event of breach. *See* Cooter, *Unity in Tort, Contract, and Property: The Model of Precaution*, 73 Calif. L. Rev. 1, 13-14 (1985).

9. This article will conventionally refer to the contract promisor as female and the promisee as male.

10. Economic cost is opportunity cost, the value of the alternative foregone as a result of the activity in question. *See* R. Posner, The Economics Of Law 6 (3d ed. 1986). The total cost of the contract to the promisor is the cost of performance, considered as opportunity cost, plus the expected liability cost of breach.

11. The parties may fail to agree to explicit remedy provisions either because the costs of the agreement process exceed the expected value of the agreement or because they prefer the remedies supplied by the law in the absence of agreement. If the law selects the remedy that would most often be selected, transaction costs are reduced.

12. R. Posner, *supra* note 5, at 82; Craswell, *supra* note 7, at 663. *But cf.* Schlag, *An Appreciative Comment on Coase's "The Problem of Social Cost:" A View from the Left*, 1986 Wis. L. Rev. 919, 929 n.39 (transaction costs may not be distinguishable from costs of production).

13. Goetz & Scott, *supra* note 6, at 972; *See* Cooter, *supra* note 8; Compare the argument in G. Calabresi, The Costs Of Accidents 26-31 (1970) that the goal of tort law is to reduce the sum of accident costs, consisting of both the losses to the injured party and the costs of avoiding accidents, consisting of investments in accident prevention and the opportunity costs of reduced production and consumption.

This does not mean that the parties will allocate contract risks simply so as to minimize outlays. "Costs" mean opportunity costs.

for a price at least equal to the costs of the remedy to her, and that a promisee would purchase only those remedies that were worth more to him than their cost to the promisor. If it can determine the relative sizes of these quantities, a court applying this method will impose any remedy the promisee would have purchased and the promisor would have sold.

This method will result in contract remedies that assign specific risks associated with breach of contract to the party who could have avoided or minimized the risk at the lower cost, the "least cost avoider."[14] The court applying this method would thus reason that a promisee must bear the risk of any loss as to which he was the least cost avoider because the parties would have contracted to allocate the risk to him.[15]

B. *Efficient Breach of Contract*

The cost avoidance theory is related to the concept of efficient breach of contract. During the period of time after entering into the contract and before performing it, a promisor may have opportunities to reallocate the resources that she would expend in performing the contract (the "performance resources") to some alternative use or purchaser. These opportunities represent a potential source of profit to the promisor if reallocation would yield a positive return after deducting contract damages and the other costs of breach.[16] Such a reallocation has been termed, somewhat misleadingly, "efficient breach,"[17] and analysts have argued that the law of contract remedies

---

14. Readings In The Economics Of Contract Law 52 (V. Goldberg ed. 1989).

15. This assumes that both parties are risk-neutral. Risk neutrality means that the person is indifferent between a certain outcome and a probabilistic one with an equal expected value. Thus, a risk neutral person would value a certainty of receiving $100 equally with a 50% chance of receiving $200 and a 50% chance of receiving 0. R. Posner, Economic Analysis of Law 11 (2d ed. 1977). Most people are not risk neutral. If the promisee is risk averse, he might be willing to shift to the promisor a risk as to which the promisee is the least cost avoider and pay the increased price the promisor would demand. But a system of remedies has no reason to assume that promisees are, as a class, more risk neutral than promisors.

16. These other costs would include costs the promisor may face in future contracts because of the effect of the breach on the promisor's reputation and the transaction costs of litigation or adjustment arising from the promisee's claim to damages for breach. This analysis will assume that the promisor has determined that the alternative use will be sufficiently profitable to cover these other costs.

17. "Efficient breach of contract" is a misleading phrase. "Efficiency" is commonly used by law and economics analysts to compare the total wealth created by alternative ways of allocating resources among parties. One allocation of resources is said to be more efficient than another if the total amount that all affected parties would be willing to pay to bring

should encourage such breaches.[18]

The opportunity for efficient breach can be analyzed as a risk as to which the promisor is the least cost avoider. To illustrate, imagine that the promisee is considering purchasing a remedy that would give him the benefit of any gain the promisor might obtain by breaching the contract and reallocating the performance resources. The cost avoidance method asks whether the value of such a remedy to the promisee would exceed its cost to the promisor. Promisors who "sell" a disgorgement[19] remedy must forego at least some opportunities for

---

it about is greater than the amount that they would pay to bring about the alternative allocation.

Because efficiency is a relative term, a breach of contract can be considered efficient only by comparing performance of the contract with some specific alternative allocation (use) of the resources that the promisor would have expended in the performance of the contract. It is the reallocation of these resources to an alternative use, not the breach per se, that may produce efficiency. Thus, it is more accurate to speak of "efficient reallocation of performance resources" than of "efficient breach."

18. "Repudiation of obligations should be encouraged where the promisor is able to profit from his default after placing his promisee in as good a position as he would have occupied had performance been rendered." Birmingham, *Breach of Contract, Damage Measures, and Economic Efficiency*, 24 RUTGERS L. REV. 273, 284 (1970); *See also* R. POSNER, *supra* note 5, at 107.

The principle of efficient breach has gained the approval of the drafters of the RESTATEMENT (SECOND) OF CONTRACTS. *Id.* Ch. 16 reporter's note. The Restatement conditions its approval upon the compensation by the breaching party of loss "calculated according to the subjective preferences of that party." *Id.* However, the law of contract damages rarely awards the full subjective value of the performance to the promisee. *See infra* notes 48-53 and accompanying text.

Critics have advised that the doctrine of efficient breach be used with caution. *See* MacNeil, *Efficient Breach of Contract: Circles in the Sky*, 68 VA. L. REV. 947 (1982) (whether different remedies will lead to efficient breach depends upon untested assumptions about transaction costs); Marschall, *Willfulness: A Crucial Factor in Choosing Remedies for Breach of Contract*, 24 ARIZ. L. REV. 733 (1982); Linzer, *On the Amorality of Contract Remedies—Efficiency, Equity, and the Second Restatement*, 81 COLUM. L. REV. 111 (1981) (arguing that using money damages instead of specific performance leads to inefficient breaches because it ignores values not easily quantified in dollars); Polinsky, *Risk Sharing Through Breach of Contract Remedies*, 12 J. LEGAL STUD. 427 (1983) (the efficiency of the expectation damages rule, compared to other damages measures, depends upon the relative risk aversion of the promisor and promisee); Farber, *Reassessing the Economic Efficiency of Compensatory Damages for Breach of Contract*, 66 VA. L. REV. 1443 (1980) (problems of detection make expectation damages undercompensatory and supercompensatory damages are often necessary to deter inefficient breach); Schiro, *Prospecting For Lost Profits in the Uniform Commercial Code: The Buyer's Dilemmas*, 52 S. CAL. L. REV. 1727, 1734-35 (1979) (arguing that inefficient breaches occur because proof requirements make expectation damages undercompensatory).

19. The term "disgorgement" is preferable to the more general term "restitution," which refers to the return of a benefit that has been conferred on the defendant by the plaintiff and the return of which is necessary to avoid unjust enrichment. *See* Farnsworth,

profitable breach that would be available to them under the expectation damages remedy. They will charge a premium for such a remedy at least equal to the value to them of the lost opportunities. The efficiency question then becomes whether promisees would value the remedy enough to pay this premium.

Cost avoidance analysis suggests that they would not. The opportunity to redirect promissory resources is generally more valuable to promisors, who, unlike promisees, are usually in the business of selling the performance resources. Promisors will have more accurate information than promisees about the likelihood that an opportunity for profitable reallocation will arise. Promisors will usually face lower costs of seeking out and selling the resources to alternative users. This cost analysis suggests that, because promisors can make more profit by redirecting performance resources than can promisees, promisors will charge more for the lost opportunities than promisees will be willing to pay and that a disgorgement remedy is inefficient.[20]

The disgorgement remedy does not, however, necessarily foreclose efficient reallocation of performance resources. A promisor who is presented with an opportunity for efficient breach could negotiate with the promisee to be released from the contractual obligation. Efficient reallocations could take place despite a disgorgement remedy.

In the frictionless fantasy worlds that economists fondly imagine, people can find each other and negotiate endlessly at no cost. In such a world, the ability to negotiate around contract remedies would make all remedies equally efficient in permitting optimum allocation of performance resources.[21] When this negotiation involves transaction costs,[22]

Your Loss or My Gain? The Dilemma of the Disgorgement Principle in Breach of Contract, 94 Yale L.J. 1339, 1342 (1985); "Disgorgement" refers to the transfer to the plaintiff of a benefit obtained by the defendant from some source other than the plaintiff and will be used in this sense throughout this article, with apologies for its unsavory alimentary connotation.

20. Promisees who anticipate having unforeseen opportunities of their own which would increase the value of performance beyond the foreseeability limitation on expectation damages can protect themselves by bargaining for liquidated damages or specific performance remedies. *See generally* Kronman, *supra* note 4; Schwartz, *The Case For Specific Performance*, 89 Yale L.J. 271 (1979). Promisees might also wish to shift risks of breach to promisors because of risk aversion.

21. Coase, *The Problem of Social Cost*, 3 J.L. & Econ. 1, 15-19 (1960); Farber, *Reassessing the Economic Efficiency of Compensatory Damages for Breach of Contract*, 66 Va. L. Rev. 1443, 1449-1450 (1980); Craswell, *supra* note 7, at 636-40 (noting that remedies will nevertheless affect the choice of contract partner and levels of precaution parties take); Schiro, *supra* note 18.

The Coase theorem suggests that, in the absence of transaction costs and wealth effects, the law's initial allocation of remedy entitlements will not affect the efficient outcome because the parties can reallocate entitlements by bargaining. Thus, if they could do so costlessly,

however, as it does in the real world, contract remedies will have differing allocative effects.[23]

The renegotiation transaction costs of a disgorgement remedy will not be insignificant. Because neither the promisor nor the promisee can deal with any other party in negotiating a buy-out of the contractual undertaking, each is motivated to try to capture the entire gain from the reallocation, secure in the knowledge that the other can deal with no one else.[24] In the absence of a market measure of their relative entitlements or claims to this gain, neither has any reason to give up any of the gain and the negotiation may never reach an agreement, even though both parties are aware that both would be better off if some division of the premium could be agreed upon. Although the parties could negotiate to an efficient outcome in a world of no transaction costs, many analysts have argued that the dead-weight cost of a "bilateral monopoly" negotiation between promisor and promisee is so high as to remove the promisor's incentive to seek out and engage in efficient reallocations.[25] Thus, a general disgorgement remedy for breach of contract seems inefficient.

---

contract parties would negotiate to the correct breach or performance decision regardless of the presence or absence of particular contract remedies. In the presence of transaction costs, such reallocations may not occur and the remedial law's initial allocation of risk will have allocative effects. Coase, *supra*, at 15.

22. Coase, *supra* note 21, at 1 (1960) (in the absence of transaction costs, parties will negotiate to efficient outcomes regardless of initial entitlements).

"Transaction costs" refers to the costs of the activities involved in engaging in a transaction, such as finding the other party, negotiating a contract, and enforcing it. Transaction costs constitute one of the more controversial elements of economic analysis. Williamson, *Transaction-Cost Economics: The Governance of Contractual Relations,* 22 J. L. & ECON. 233 (1979); Goldberg, *Production Functions, Transaction Costs, and the New Institutionalism,* ISSUES IN CONTEMPORARY MICROECONOMICS 395-402 (G. Feiwel ed. 1984), *reprinted in* READINGS IN THE ECONOMICS OF CONTRACT LAW, 21-23 (arguing that if "transaction costs" merely refers to anything that prevents an efficient outcome, the concept is a useless tautology).

23. *See supra* note 22.

24. R. POSNER, *supra* note 5, at 107. Economists refer to this behavior as "rent seeking." *See* Goldberg, *supra* note 22, at 50, 70, 71. One problem is that the promisor may be reluctant to disclose the existence of her opportunity for reallocation to the promisee for fear that he will take advantage of it. Because of this reluctance and the mistrust of bargaining, the promisee will have to guess at how much potential profit the promisor could make from reallocation. This uncertainty might lead the promisee to make unrealistic demands.

25. *See, e.g.*, R. POSNER, *supra* note 5, at 117-18 (discussing specific performance); Barton, *The Economic Basis for Damages for Breach of Contract*, 1 J. LEGAL STUD. 277 (1972); Goetz & Scott, *supra* note 5, at 982-83 (discussing opportunism in renegotiation).

Some dissenters from this view contend that the transaction costs of negotiation do not exceed the transaction costs of "breach first, talk afterward" that are encouraged by the expectation damages remedy. MacNeil, *supra* note 18, at 968; *See also id.* at 959-60,

C. *The Efficiency of Traditional Disgorgement Remedies For Breach of Contract*

Despite the arguable inefficiency of a general remedy of disgorgement, restitutionary remedies[26] are common in various "failed contract" situations.[27] Parties to valid contracts are often awarded restitution of

---

n.42 (noting that the expense of bilateral monopoly negotiations depend on several factors, including the relations between the parties); Kronman, *supra* note 4, at 353 (negotiations for a voluntary transfer of contract rights between existing parties to a contract avoid many costs, such as search costs, that might attend other forms of negotiation); Farber, *supra* note 18, at 1453-54 (transaction costs requiring an alternative purchaser to negotiate with the original buyer probably less than the transaction costs of breach and settlement or litigation); Schwartz, *supra* note 20, at 284-91.

26. *See supra* note 19 on the distinction between restitution and disgorgement. While the expectation damages remedy seeks to achieve the goal of compensation (i.e., making the plaintiff whole) the remedy of restitution seeks to achieve the goal of preventing unjust enrichment, compelling the defendant to disgorge profit wrongfully obtained. A. Corbin, *supra* note 4, at 1104 (defining "restitution"); G. Palmer, Law of Restitution § 1.1 (1978); Restatement (Second) of Contracts § 370. *See also* Barton, *supra* note 25.

27. E. Farnsworth, Contracts § 12.19 (1982) (which lists several such situations).

One must distinguish between restitution as a remedy for quasi-contractual claims and restitution as a remedy for breach of contract. A plaintiff may recover restitution under the quasi-contract doctrine in the absence of contractual promise. Thus, the law will "imply" a contract to pay restitution when the plaintiff mistakenly confers a benefit on the plaintiff. Restatement (Second) of Restitution (Tent. Draft No. 1 April 5, 1983) sec. 8. The law will also provide a restitutionary remedy in cases of promisory estoppel, in which the defendant's promise induced reasonable detrimental reliance on the part of the plaintiff. Restatement (Second) of Contracts § 90.

Often the restitutionary remedy is available in various forms of "failed contract" situations, in which for some reason a contemplated contract has not been formed and even where the defendant has not been enriched. Dawson, "Restitution Without Enrichment," 61 Boston U.L.R. 563 (1981); *See e.g.*, Kearns v. Andree, 107 Conn 181, 139 A. 695 (1928) (restitution of seller's costs of renovating property at buyer's request although sale contract was unenforceable because of vagueness); Farash v. Sykes Datatronics, Inc., 59 N.Y.2d 917, 452 N.E. 2d 1245 (1983) (awarding restitution of lessor's costs of renovating leasehold at lessee's request although oral lease was unenforceable under Statute of Frauds). *Cf.* Boone v. Coe, 153 Ky. 233, 154 S.W. 900 (Ct. App. Ky. 1913) (denying restitution where plaintiffs' expense in preparing to perform a contract unenforceable under the Statute of Frauds did not benefit defendant).

Restitution is available as a remedy to require return of benefits conferred by a non-breaching party upon a breaching party in some cases of breach of contract. Restatement (Second) of Contracts § 371; E. Farnsworth, Contracts §§ 12.19, 20 (1982); Friedman, *Restitution of Benefits Obtained Through the Appropriation of Property or the Commission of a Wrong,* 80 Colum. L. Rev. 504, 513. A non-breaching party who has rendered partial performance may recover the value of that performance even if that recovery exceeds the party's expectation interest. 12 S. Williston, A Treatise on the Law of Contracts § 1485 (W. Jaeger ed. 1970) (recovery should be undiminished by any loss promisee would have incurred on performance); United States v. Algernon Blair, Inc., 479 F.2d 638 (4th Cir. 1973) (nonbreaching subcontractor could recover from breaching prime contractor in

benefits which they have conferred upon other parties.[28] However restitution as a remedy for breach of contract is limited to benefits conferred upon the defendant by the plaintiff,[29] and only rarely is a party awarded the remedy of disgorgement of profits the breacher obtained from some other source, as she would by reallocating performance resources to another use.[30] In most situations in which disgorgement is permitted, one can argue that it is not inconsistent with the goal of efficiency.

Thus, what appears to be a disgorgement remedy may simply be the most convenient method of measuring the plaintiff's expectation interest.[31] For example, a disgorgement remedy is common in contracts for the sale of real property. Sellers who breach a contract to sell land by selling it to a third party have been required to pay the non-breaching buyer the amount by which the sale price to the third party exceeded the contract price.[32]

---

quantum meruit for labor and materials furnished even though the contract would have been a losing contract for subcontractor). *See* E. FARNSWORTH *supra,* § 12.20.

Restitution is denied, however, where the plaintiff has fully performed and the only breach by the defendant is payment of the contract price. Restatement (Second) of Contracts § 371; Oliver v. Campbell, 43 Cal. 2d 298, 273 P.2d 15 (1954 (*quantum meruit* denied for attorney who had fully performed contract for legal services); Fay, Spofford & Thorndike, Inc. v. Massachusetts Port Authority, 7 Mass. App. Ct. 336, 387 N.E.2d 206 (1979) (justifying the rule as avoiding unnecessary valuation problems).

28. One example of disgorgement as a remedy for breach is the requirement that a promisor who is guilty of total breach or repudiation must return the full amount of any contract payments received, even if they exceeded the promisee's expectation damages. E. FARNSWORTH, *supra* note 27, § 12.20; RESTATEMENT (SECOND) OF CONTRACTS § 373 comment d. In Bush v. Canfield, 2 Conn. 485 (1818), the buyer paid seller $5,000 in advance for 2,000 barrels of flour at $7 per barrel. Seller failed to deliver, although at the time of delivery the price had dropped to $5.50. Buyer would have lost $3,000 on the sale. The court refused to deduct the $3,000 from the damages and ordered the seller to return the entire price paid. The Uniform Commercial Code goes farther, giving the buyer a right to the return of any purchase price paid not only upon the seller's repudiation or failure to deliver but also upon the buyer's rejection or revocation of acceptance, with respect to any goods involved and regardless of whether the contract was a losing one for the buyer. U.C.C. § 2-711(1). *see also* § 2A-508(1)(b) for similar rule applying to leased goods.

29. RESTATEMENT (SECOND) OF CONTRACTS § 370. The exceptions to this limitation are not considered in this article. *See* J. CALAMARI & J. PERILLO, CONTRACTS 652-53 (3d ed. 1987).

30. G. PALMER, *supra* note 26, § 4.9 at 437 (generally "mere breach of contract will not make the defendant accountable for benefits thereby obtained, whether through dealings with a third person or otherwise."). Commentators have argued for limited extensions of the right to restitution of the breacher's benefits. *See* Farnsworth, *supra* note 19, at 1384-86 (arguing for extension of the disgorgement remedy to "abuse of contract" cases); Friedmann, *supra* note 27, 516-27 (arguing for restitution where the breach has deprived the promisee of a property or "quasi-property" interest).

31. Farnsworth, *supra* note 19, at 1360; Dawson, *Restitution or Damages*, 20 OHIO ST. L.J. 175, 179 (1959).

32. *See* cases cited in Friedmann, *supra* note 27, at 516; G. PALMER, *supra* note 26,

Disgorgement in such cases can be seen as a measurement of the plaintiff's expectation interest if one assumes that the plaintiff could have sold the land to the alternative purchaser.[33] The fixed location of real property and the existence of property records make it likely that the alternative purchaser could have found and negotiated with the plaintiff as easily as with the defendant. It is reasonable for the law to presume that the buyer could have sold it to the alternative purchaser at the price received by the seller, which is, therefore, selected as a convenient measure of the expected value of performance to the buyer.

Similar use of disgorgement to measure the promisee's expectation occurred in *Laurin v. DeCarolis Constr. Co., Inc.*[34] Defendant seller contracted to sell land to the plaintiff. After the contract but before the closing, the seller removed and sold gravel from the land. Even though testimony indicated that removal of the gravel had not diminished the market value of the land, the plaintiff recovered the seller's net profits on the gravel, the court reasoning that the plaintiff could have obtained those profits had the seller not breached.[35]

Disgorgement also serves as a measure of plaintiff's loss in the award of profits made on sales in breach of a covenant not to compete.[36] Proof of the plaintiff's actual lost profits in such cases is highly speculative because of the difficulty in proving that plaintiff would have gotten the business the defendant enjoyed. The convenient assumption is made that the plaintiff would have profited by as much as the defendant did.

In these instances, disgorgement is probably the best measure of the expectation damages plaintiff suffered. Despite inhibiting promisors from efficient breach, the remedy is not thought to be inefficient, in part because the promisors have no cost advantage over the promisee in reallocating the performance resources to the alternate use.

In addition to the traditional uses of disgorgement outlined above, Professor Farnsworth has suggested the remedy of disgorgement for what he calls "abuse of contract."[37] Farnswoth argues that disgorgement is

---

at 438; Timko v. Useful Homes Corp., 114 N.J. Eq. 433, 168 A. 824 (1933). In such cases, courts sometimes characterize the non-breaching buyer as having acquired a beneficial ownership in the land by entering the sales contract. The seller's subsequent sale of the land to a third party is seen as a conversion of this beneficial interest. Another approach has been to attribute the sale to a fictional agency relationship between the buyer and seller.

33. Friedmann, *supra* note 27.

34. 372 Mass. 688, 363 N.E.2d 675 (1977).

35. The court in *Laurin* rejected a claim for conversion, on grounds that plaintiff was not entitled to possession at the time of the removal of the gravel. *Id.* at 689, 363 N.E.2d at 677.

36. Farnsworth, supra note 19, at 1367.

37. *Id.*

appropriate when the breach denies the promisee an opportunity to obtain substitute performance. He considers the classic situation of the contractor who secretly substitutes cheaper materials in the construction of a building under a contract calling for more expensive materials.[38] By the time the breach is discovered, the expense of tearing down and rebuilding the structure greatly exceeds the diminution in market value caused by the breach. In such cases, courts usually limit the promisee's recovery to the diminution in value of the structure.[39] Because the promisee can no longer purchase the performance he desired, Farnsworth argues that the promisor should also be required to disgorge the savings she realized in using the cheaper materials.

Cost avoidance analysis suggests that disgorgement would be efficient in such circumstances. Assume that *ex ante* a contractor knows that the project will offer her opportunities to "cheat" by substituting cheaper materials. These opportunities will reduce her costs and increase the profit she can make. If she successfully escapes detection by the promisee until re-construction becomes unfeasible, then under the traditional judicial rule, she can capture the difference between these cost savings and the diminution in market value caused by the substitution, if any. It is important to note that this figure will be at most equal to the cost savings, and could be less. The expected value of this capture will be taken into account in pricing the job, reducing the contractor's costs by an equal amount.

Now assume that Farnsworth's proposed remedy of disgorgement is put into place. The promisor's incentive to cheat will be removed because the promisor will have to give back any cost savings otherwise obtainable by substitution. The contractor's opportunity cost of the contract will rise by the amount of the forgone cheating premium, and the price of the contract to the promisee will rise by the same amount.

Whether the disgorgement remedy is efficient in this situation depends upon whether the promisee would be willing to pay this increased price. This question turns on the amount by which the promisee values the use of the specified materials over the inferior materials. If this amount exceeds the cheating premium, Farnsworth's proposed rule is efficient.

There is good reason to believe that the promisee values the superior materials more highly than the inferior materials by more than their

38. The classic chestnut in this field is Jacob & Youngs, Inc. v. Kent, 230 N.Y. 239, 129 N.E. 889 (1921) in which Judge Cardozo held that where a contractor inadvertantly substituted pipe of the wrong brand, the owner was entitled to recover only the reduction in value of the building, rather than the cost of tearing out and replacing the pipe, on grounds that the breach was "insubstantial."

39. E. FARNSWORTH, *supra* note 26, § 12.13, at 868. The rule is otherwise if the cost of repairs is insubstantial. If the breach is willful, the finding of substantial performance may be questionable. *Id.* at 593-94.

price difference: He must value them by at least their price difference (otherwise he would have specified the cheaper materials)[40] and he might value them by more. Thus, while the cheating premium is less than or equal to the price difference, the value of the remedy to the promisee is greater than or equal to the price difference. On these assumptions, Farnsworth's remedy is likely to be more valuable to promisees than it is expensive to promisors.

Disgorgement is also an efficient remedy in contracts in which the promisor is financially unable to respond in damages. In such situations, disgorgement can discourage inefficient, opportunistic behavior that the expectation damages remedy would encourage. An example will illustrate: Assume that in March, Seller contracts to sell Buyer widgets for delivery in June. Seller's expected profit is $1 per widget. In May, the market having risen, Alternative Purchaser offers Seller the then market price for the widgets, which is $15 over the contract price. Widgets are relatively fungible but the market is highly volatile. Seller knows that there is a 90% chance that, at the time Buyer must cover,[41] the market will have risen to $50 over the contract price and a 10% chance that it will have fallen to $3 over the contract price.

If Seller will be financially able to pay the Buyer's full claim for damages, she will reject the offer: 90% of the time she will receive $16 in gross profit but will have to pay Buyer $50 damages, leaving her $34 in the hole. Ten percent of the time she will have to pay the Buyer only $3, leaving her a net profit of $12. The combined value of these chances, or the expected value of the breach, is minus $29.40,[42] which is less than the certain profit of $1 that performance would yield. In other words, the breach is inefficient.

However, if the Seller is judgment-proof, so that she cannot pay Buyer's claim for damages, the gamble is a good one. The effect of market rise is simply to wipe out the chance of any profit, not to subject the Seller to a loss of $34. The effect of a market fall still gives her a $12 profit. The expected value of breach is $1.20,[43] which exceeds the $1 profit available if Seller performs.

Because the judgment-proof seller can speculate with less risk than the financially responsible seller, an expectation damages remedy en-

---

40. This analysis assumes that the promisee knows of the price and the value of the substituted materials. If not, it cannot be assumed that the promisee would not have specified them in the original contract.

41. It is assumed that the seller can sell to the Alternative Purchaser without repudiating the contract and thus that the buyer's remedy, by cover or market formula, will be measured as of a later time.

42. $(0.9 \times -34) + (0.1 \times 12) = -29.40$.

43. $(0.9 \times 0) + (0.1 \times 12) = 1.20$.

courages judgment-proof sellers to engage in inefficient breaches. Under a general disgorgement remedy for breach of contract, the judgment-proof seller will not engage in inefficient breaches of this sort because the remedy removes any opportunity for gain. Such a remedy tends to reduce the transacting costs related to insuring against the credit risks of doing business with judgment-proof promisors. In summary, although cost analysis suggests that disgorgement would not be efficient as a generally available remedy for breach of contract, the remedy would appear to be efficient in some situations. First, disgorgement may be efficient when expectation damages are difficult to calculate and are likely to approximate the disgorgement measure. Second, disgorgement can be more efficient than expectation damages when the promisor's inability to recover the full amount of expectation loss creates incentives for the promisor to engage in certain types of inefficient, opportunistic behavior.

D. *The Efficiency of Disgorgement as a Remedy for Conversion*

Although inefficient as a general remedy for breach of contract, disgorgement is a common remedy for the tort of conversion. Traditional restitutionary remedies permit a plaintiff suing a convertor to "waive the tort" and to require the defendant to disgorge all profits made as a result of the conversion, even if they exceeded the plaintiff's losses.[44]

Some of the efficiency arguments made against disgorgement as a contract remedy seem equally applicable to disgorgement as a tort remedy. It is obvious that conversion can be just as transactionally efficient as a breach of contract. The convertor might make more profit from the plaintiff's property than the plaintiff would lose from the unlawful use. Thus, requiring disgorgement of such profits weakens incentives for convertors to seek out and enter into efficient "reallocations."

Yet, contemporary law and economics literature does not propound a doctrine of "efficient conversion." A different efficiency argument is marshalled to support the law's refusal to countenance involuntary transactions of this sort, one that goes something like this: Even if the converted good is traded in a recognized market, a court cannot be sure of the subjective value of the good to the owner and so cannot be certain of the efficiency of reallocating the good to the convertor. A judicial determination of the value of the converted good to its owner incurs this cost of uncertainty as well as significant transaction costs. Because the convertor can obtain the converted good from the market at much lower transaction costs, the law requires her to obtain the goods

---

44. *See* Friedmann, *supra* note 27; Dawson, *supra* note 27; Farnsworth, *supra* note 19, at 1361.

in a voluntary exchange to insure that her use will indeed be allocatively efficient.[45]

E. Disgorgement Reconsidered

The analysis of conversion has some application to the previously stated efficiency justification for rejecting the disgorgement remedy for breach of contract. A contract breacher, like a convertor, makes a unilateral decision to reallocate resources and relies on a post-breach judicial decision to fix the amount of the promisee's damages. Our previous efficiency analysis of the disgorgement remedy did not consider the transaction costs of a judicial determination of contract damages because of the implicit assumption that expectation damages give the promisee the equivalent of full performance.[46]

If expectation damages fulfilled their theoretical purposes, promisees would be indifferent as to whether contractual promises are performed or damages are awarded.[47] In practice, however, promisees are not indifferent to performance because awards of damages under the ex-

---

45. Calabresi & Melamed, *Property Rules, Liability Rules and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972). Analogizing the efficiency of rules against conversion to the efficiency of rules against theft, Landes and Posner argue that depriving the convertor of his profit will have the efficient effect of compelling him to resort to the marketplace whenever possible. *See* W. LANDES & R. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 153-58 (1987). They also argue that failure to protect property interests would lead owners to over-invest in security measures to prevent unlawful takings. *Id.* This argument seems questionable. If owners of converted property or beneficiaries of looted trusts are awarded full compensatory damages, they would be indifferent whether the torts were committed or not. Because they would experience no economic loss as a result of the wrong, they would not spend excessive amounts on security. If remedies are undercompensatory, of course, this argument stands up, but so would an analogous argument applied to remedies for breach of contract.

46. Professor MacNeil offers an example of the arbitrariness of distinguishing conversion from breach of contract on the basis of "property" interests by an example of a seller who "reclaims" goods that have been shipped to the buyer's warehouse in order to engage in efficient breach. The buyer would be entitled to disgorgement for conversion, yet in many complex commercial transactions, "there is no *a priori* basis for selecting any particular time or event for determining that the transaction is closed, that new property rights are now established, and that with them a new, efficiency-neutral status quo has been achieved." MacNeil, *supra* note 18, at 963-64.

47. R. POSNER, *supra* note 5, at 107. *But see* U.C.C. § 2-609 comment 1:

> ". . .the essential purpose of a contract between commercial men is actual performance and they do not bargain merely for a promise, or for a promise plus the right to win a lawsuit and . . . a continuing sense of reliance and security that the promised performance will be forthcoming when due is an important feature of the bargain."

pectation doctrine rarely give a promisee the complete value of his loss.[48] Many forms of a promisee's loss are not recoverable: losses whose existence is uncertain;[49] losses that were not foreseeable at the time the contract was formed;[50] losses that the promisee could have avoided or mitigated;[51] losses from emotional disturbance;[52] and losses from the expenses of contract enforcement.[53] Whether or not these limitations on recovery are justified by considerations of rule efficiency,[54] the result is

---

48. *See* Schiro, *supra* note 18, (arguing that remedies under the Uniform Commercial Code systematically undercompensate plaintiffs for lost profits and therefore encourage inefficient breaches); Farber, *supra* note 18, (arguing that, because of litigation costs and detection costs, "supercompensatory" damages are more efficient than expectation damages); Schwarz, *supra* note 20, at 276-77 (damage awards are undercompensatory); Sebert, *Punitive and Nonpecuniary Damages in Actions Based Upon Contract: Toward Achieving the Objective of Full Compensation*, 33 UCLA L. Rev. 1565 (1986). This has suggested to many commentators that expectation damages should be made more compensatory.

It might also be argued that expectation damages are insufficient to prevent inefficient breaches because they are not awarded to non-contracting parties at the time of formation of the contract, but nevertheless must be "counted" in determining the transactional and rule efficiency of breach, because they are part of the "aggregate willingness to pay" on which the criterion is premised.

49. Restatement (Second) Of Contracts § 352. *See* Harry Rubin & Sons, Inc. v. Consolidated Pipe Co. of America, 396 Pa. 506, 153 A.2d 472 (1959) (under the U.C.C., refusing to award damages for loss of goodwill as too speculative).

50. *E.g.*, Restatement (Second) Of Contracts § 351; Hadley v. Baxendale, 9 Ex. 341, 156 Eng. Rep. 145 (Ex. 1854); Globe Refining Co. v. Landa Cotton Oil Co., 190 U.S. 540 (1903); Keystone Diesel Engine Co. v. Irwin, 411 Pa. 222, 191 A.2d 376 (1963) (refusing to award lost profits). *See also* Restatement (Second) of Contracts § 350.

51. Restatement (Second) of Contracts § 350; Rockingham County v. Luten Bridge Co., 35 F.2d 301 (4th Cir. 1929).

52. Restatement (Second) Of Contracts § 353. Emotional distress reduces aggregate human satisfaction if people would pay to avoid it. *See supra* note 8. If the promisor is forced to negotiate with the promisee before the breach, the value of the emotional distress of breach can enter into the negotiation and an otherwise inefficient breach can be avoided. On the other hand, failing to award damages for emotional distress makes sense under a regime of expectation damages because it encourages parties who have idiosyncratic attachments to contract performance to disclose these and secure additional security for performance from the promisors.

53. Thus, under the "American Rule" in the absence of a contractual agreement, the attorney's fees expanded in enforcing a contract remedy are not recoverable. *Alyeska Pipeline Service Co. v. Wilderness Society*, 421 U.S. 240, 247 (1975). *See* Restatement (Second) of Contracts § 356 comment d. It has been argued that, because of the expenses of enforcement, the non-breaching party should always be willing to accept less than the expectation damages from the breaching party. *See* Leff, *Injury, Ignorance and Spite—The Dynamics of Coercive Collection*, 80 Yale L.J. 1 (1970).

54. An efficiency analysis of the limitations on expectation damages is beyond the scope of this article. *See, e.g.*, R. Posner, *supra* note 5, at 114-15 (foreseeability requirement efficiently requires promisees to disclose potential losses in advance, permitting promisors to take appropriate precautions); Cooter, *supra* note 8 (contract damages limitations create

that contract damages are undercompensatory.[55] Expectation damages do not measure the promisee's willingness to pay for performance at the time of breach. The limitations on expectation damages will lead to transactionally inefficient breaches.[56] Because of problems of subjective valuation, the promisee's consent to a reallocation of performance resources is the only certain proof of the efficiency of a breach of contract.

The problematic nature of the judicially determined expectation damages remedy raises the issue considered in the analysis of "efficient conversion." If the cost of uncertainty in a post-conversion determination of damages exceeds the transaction costs of requiring potential tortfeasors to resort to the market, it is possible that the cost of uncertainty in post-breach determination of damages likewise exceeds the transaction costs of requiring a promisor to negotiate with the promisee before breaching. Requiring a voluntary exchange between the promisor and promisee before breach would provide assurance that the breach will be truly efficient, just as would requiring a voluntary exchange between the convertor and her victim.[57] But this gain in certainty would be purchased at the expense of any lost opportunities for efficient breach that were frustrated by the costs of negotiation.

The problem reduces to this: Are the transaction costs of a bilateral monopoly negotiation between the two contract parties "relatively high" so that the court is justified establishing a liability rule, permitting unilateral breach on condition that expectation damages be paid, or "relatively low" so that the court is justified in requiring recourse to the market to avoid the risk of an inaccurate determination of damages?[58] To date most economic analysis has answered "relatively high," and has supported a liability rule rather than a property rule[59] for the

---

efficient incentives to minimize losses); Goetz & Scott, *supra* note 6 (efficiency of mitigation requirement); Goetz & Scott, *Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach,* 77 Colum L. Rev. 554, 568-74 (1977).

55. *See supra* note 18. Even when the promisee is the least cost avoider of a risk of loss due to breach, the promisor's failure to take that risk into account in deciding to breach will lead to transactionally inefficient breaches whenever the promisor would be willing pay more for performance than the alternative user at the time of breach.

56. The theory of efficient breach posits that the promisor will breach whenever she can show a profit after paying the promisee's expectation damages. But in such cases, breach will be inefficient when the promisor's gain from breach does not exceed the amount by which expectation damages fall short of actual value of performance to the promisee.

57. *See* W. Landes & R. Posner, *supra* note 45.

58. *See* Schlag, *supra* note 12, at 930-31 n.43 (asking what a court should do in situations in which the transaction costs of trading in entitlements are "middle.").

59. *See* the discussion in Kronman, *supra* note 4, citing Calabresi & Melamed, *supra* note 45.

promisee's contract interest. The breacher is free to breach without obtaining the promisee's consent, but must pay damages.

That this conclusion is correct is not at all obvious, but it is not within the scope of this Article to test it, or even to speculate about how that might be done. Instead, two other facets of the disgorgement remedy that might bear on its efficiency will be examined. First, even if one accepts that the promisee's interest in performance is protected by a liability rule, should the promisee's interest in payment of damages for breach be protected by a liability rule or a property rule? Second, is a property rule, a disgorgement remedy, practically feasible as a remedy for protection of either interest?

III. DAMAGES FOR FAILURE TO PAY DAMAGES

Although a party in breach of contract becomes liable to the non-breaching party for payment of damages, the breaching party may, and usually does, withhold payment of damages until ordered to do so by a court. General use of the term "efficient breach of contract" has obscured the distinction between two separate breaches: the breach of the duty to perform and the breach of the duty to pay damages for non-performance. A breacher who fails to pay the full amount of the promisee's damages at the time of breach[60] can obtain economic gain through the use of the money she should have paid. The promisee can incur economic loss because of the delay in receiving payment.

The literature of efficient breach has not addressed the possibility that failure to pay contract damages when due might itself be efficient if the promisor can earn more on the money than the promisee would

---

("[A] right or entitlement is protected by a property rule when it can be appropriated by a non-owner only if he first purchases permission to do so from the owner of the right. When a right is protected by a rule of this sort, one who appropriates it without the owner's permission will always be subject to a special sanction—typically, a fine or imprisonment. If a right is protected by a liability rule, in contrast, a non-owner who unilaterally appropriates it need only compensate the owner, after the taking, for any loss the owner suffers. The compensatory amount which a non-owner must pay for taking a right protected by a liability rule is set by a representative of the state rather than by the owner of the right in a voluntary transaction between owner and taker.").

The expectation damages rule is a liability rule and the disgorgement rule is a property rule.

60. Much will depend upon the time when breach occurs. Breach normally occurs at the time of performance. However, breach may occur before the time of performance if the promisor repudiates the contract. Hochster v. De la Tour, 2 El. & Bl. 678 (1853); Missouri Furnace Co. v. Cochran, 8 F. 463 (W.D. Pa. 1881); U.C.C. § 2-610 (permitting non-breaching party to resort to any remedy for breach upon promisor's repudiation of contract).

lose by not receiving it. The above discussion suggests two possible analyses of the efficiency of withholding damages, corresponding to the liability and property rules, or expectation damages and disgorgement. A liability rule/expectation damages approach would promote efficient withholding of damages and would evaluate the rule efficiency of interest remedies by their tendency to encourage or deter efficient withholdings. A property rule/disgorgement analysis would reject efficient withholding of damages and would evaluate interest remedies by their tendency to encourage the breaching party to borrow money in the financial market rather than to "borrow" damages from the plaintiff.

A. Ex Ante Cost Avoidance Analysis of Prejudgment Interest

The law recognizes a claim for interest, damages for delay in the payment of money due, as a remedy for breach of contract.[61] As with any other contract remedy, a promisee's right to recover interest accruing before judgment will affect allocative efficiency. The efficiency of the disgorgement remedy of prejudgment interest can be analyzed by the cost avoidance method: Would the parties be likely to give the promisee a disgorgement remedy for failure to pay contract damages immediately upon breach? They will do so only if the benefit of the remedy to the promisee/plaintiff exceeds the opportunity cost of such a remedy to the promisor/breacher. These values can be analyzed at the time of contracting, *ex ante*, and after breach, *ex post*. The parties would presumably prefer the remedy that maximized the joint value of the contract.

Under a compensation theory of prejudgment interest the promisor is assigned the power to withhold contract damages. Under a disgorgement theory of prejudgment interest the promisee is granted the right to immediate payment of contract damages. There is no *ex ante* reason to suppose that either promisors or promisees will have, as a class, superior investment opportunities at the time of breach.[62] Thus, there appears to be no reason to believe that relative opportunity costs alone would lead to an *ex ante* assignment of the risk of delayed payment of contract damages to either party.

---

61. *See* Restatement (Second of Contracts § 354. Although parties may agree to the payment of interest as compensation for the use of money, such agreements are not discussed in this article, which concerns only judicial remedies for the non-payment of contract money damages. Such agreements may provide for interest to be paid either before or after maturity of the obligation. See 5 A. Corbin, *Supra* note 4, § 1045; D. Dobbs, Handbook On The Law Of Remedies 3.5 (1973).

62. Promisees will often have need of funds to effect cover or other mitigation of the injury caused by the breach. Promisors may, however, have need of the funds to effect the alternative, more profitable, performance. It is impossible to say which of these two forms of investment has the greater expected return in the abstract.

B. Post-Breach Cost Avoidance Analysis of Prejudgment Interest

The method chosen for the award of prejudgment interest will affect the post-breach behavior of both the breaching promisor and the promisee, who will have the relation of debtor and creditor.[63] Both debtors and creditors have borrowing costs[64] as well as opportunities to obtain returns on money they invest. It is assumed that parties will tend to borrow whenever their investment returns exceed their borrowing costs.[65] The efficient remedy would "assign" the post-breach damage payment to the party with the superior post-breach investment opportunity. Only the compensation method tends to have this result, as is shown by the following analysis.

1. *Compensation Method.* The compensation method compensates the creditor for the deprivation of the use-value of the money until payment. The measure of this loss is either the creditor's cost of cover—the interest that the creditor must pay to borrow the funds elsewhere—or the loss of profit on investment that the creditor was unable to make because of the lack of the money. The principle of mitigation of damages[66] would deny the creditor any losses that could have been avoided had the creditor covered by obtaining funds elsewhere at a rate that is less than his investment return. If the creditor cannot cover, his loss is the forgone investment return, if any.

If a compensation remedy is used, debtors will tend to withhold payment whenever they will earn more from the withheld money than they will ultimately have to pay to the creditors. The compensation remedy would award creditors the lesser of their borrowing expense in obtaining "covering" loans or, if profitable cover is impossible, their forgone investment returns. Debtors will withhold payment only when both their borrowing costs and their investment returns are higher than their creditors' losses. Such debtors can, in effect, borrow at below market rates from their creditors and would lose more by paying the debt than they would ultimately owe in damages measured by the creditors' losses.

---

63. The following discussion assumes that the parties are risk-neutral.

64. For the purposes of this discussion, "borrowing costs" include returns on existing investments that must be foregone if the investment is liquidated to obtain cash for payment of a debt or to make another investment. In such cases, the party in question is, in effect, borrowing from himself.

65. As usual, this analysis assumes that the parties are risk-neutral. Risk-averse parties might not borrow to invest. This analysis also ignores the possibility of consumption rather than investment of the indebtedness. Consumption can be seen as an investment that has a value known to the consumer.

66. RESTATEMENT (SECOND) OF CONTRACTS § 350.

If either the debtors' borrowing costs or investment returns are less than the losses of their creditors, the compensation method would induce debtors to pay promptly. If debtors' borrowing costs are less than the damages they owe, they would be better off borrowing and paying damages. If debtors' investment returns are less than the damages they will owe, any gains they would realize by withholding the funds would be exceeded by their liability for their creditors' losses.

As can be seen, under the compensation method the money owed will tend to move toward the party with the highest investment opportunities, whether or not that party has the higher borrowing costs. This apparently efficient result is not achieved by the disgorgement method, however.

2. *Disgorgement Method.* The disgorgement method measures the benefit to the debtor, who enjoys the use-value of the money until payment. The measure of this benefit is either the borrowing costs saved because the debtor did not have to borrow the money elsewhere or the return on any investment that the debtor made with the money.

If the disgorgement method is used, debtors with superior investment opportunities will tend to pay the debts regardless of their borrowing costs or the losses of their creditors.[67] They have no motive to withhold payment because any profit they may make will be captured by the creditors. This capture might sweep in some profits resulting from the debtors' own skill or efforts in connection with the investments.

The result under the disgorgement method is more ambiguous if debtors with no advantageous investment opportunities have higher borrowing costs than the damages they will owe. If such debtors must borrow to pay the debt, then their "benefit" from non-payment is their borrowing cost. Even if they must ultimately disgorge this benefit to their creditors, they have no incentive to choose repayment over withholding. In such cases, and in cases in which debtors can pay the debt without borrowing, the following analysis applies.

If the disgorgement method is the creditors' exclusive remedy, so that they can recover only their debtors' gain and not their own losses, debtors with inferior investment opportunities have nothing to lose by withholding and nothing to gain by repayment. Such debtors might strategically withhold payment. Their creditors, facing losses larger than their potential recovery, may offer a discount for prompt payment.

If the disgorgement method is supplemented by the compensation method, so that creditors can recover the greater of their losses or their

---

67. Such debtors may negotiate with their creditors to obtain loans at rates that are less than the debtors' borrowing costs but greater than their creditors' investment returns. This does not, however, affect the analysis.

debtors' gains, such debtors will pay promptly in all cases unless they are judgment proof. They will have nothing to gain and everything to lose by withholding payment.

In summary, under the "pure" disgorgement method, unsupplemented by a compensation remedy, the money will, without negotiated transfers, tend to remain with or gravitate to the party with the poorest investment opportunity. Under a combination compensation and disgorgement method, the money will move toward the creditor regardless of the parties' investment opportunities.

C. The Effect of Borrowing Rates

So long as the effects of different borrowing rates are ignored, the compensation method of prejudgment interest appears to be more efficient than the disgorgement method in allocating money to the party whose investment opportunities are the highest. But this analysis ignores the factors that determine the "borrowing rate" of each party. Variations in this rate are almost entirely due to the riskiness of the loan, that is, the creditworthiness of the party.[68] If the contract creditor is required to mitigate his losses by obtaining funds elsewhere, his borrowing costs will be the measure of his losses.[69] The compensation method permits the debtor, in effect, to utilize her creditor's lower borrowing costs, or stronger credit rating, as her own expense of borrowing. The debtor obtains money at a cost below the rate that she would be charged in the money market. The creditor, however, is not being compensated for the risk associated with the debtor's use of the money, risk that would be reflected in the debtor's borrowing rate.

Like a potential convertor, a contract debtor can obtain the withheld good (money) in a market transaction. The debtor can enter the market for the withheld good at low transaction costs. Lenders in established credit markets probably face lower costs of extending credit to a debtor[70]

---

68. R. Hamilton, Fundamentals of Modern Business § 1.16 (1989); *See* R. Posner, *supra* note 5, at 180, 368. The other elements of the interest rate are common to all borrowers and include the time value of money and the normal rate of return.

69. A debtor should not be permitted to count her creditor's actual borrowings and investments as "covering" transactions if they would have been economically rational in the absence of the withholding. One might expect a risk-neutral creditor to borrow and invest whenever his investment opportunities exceed his borrowing costs. By analogy to the theory of the lost volume seller, a creditor's "covering" loan and investment should not reduce the debtor's liability if it would have been made in the absence of the withholding. To show damage from the withholding, however, a creditor should have to show that he had exhausted his ability to borrow and still missed favorable investment opportunities.

70. This would include the costs of obtaining and evaluating information relevant to the riskiness of the loan and of insuring against unavoidable credit risks by loss spreading. *See* R. Posner, *supra* note 5.

than does her contract creditor. The contract creditor is, therefore, not the least cost avoider of the costs of the debtor's credit risks and should not be assigned that risk by the compensation method of prejudgment interest.[71] In addition, a post-breach judicial determination of a creditor's compensatory damages presents a special risk of inaccuracy because his forgone investment opportunities might be, or appear to be, speculative or convenient afterthoughts. Despite its apparent allocative inefficiency, the disgorgement measure of prejudgment interest is more efficient than the compensation method for reasons similar to those pertinent to the conversion remedy.

IV. A CASE STUDY IN DAMAGES AS INTEREST

We now descend from the crystalline forms of economic theory to the woolly world of law as it is. In practice, remedies law rarely employs either the compensation method or the disgorgement method as a measure of prejudgment interest. The requirements of mitigation and foreseeability limit inclusion of consequential damages in prejudgment interest.[72] Prejudgment interest is awarded, if at all, at a statutory rate.[73] Such statutes have provided for interest at a fixed rate to be awarded on amounts found to be due for breach of a duty to pay liquidated sums due under contracts, notes, and bills; for money obtained by fraud; and in other situations.[74] Statutory interest may roughly approximate an award under the compensation method, which is intended to compensate the creditor for the loss of use of the money.[75] Interest is measured by the disgorgement method only when the money was obtained by breach of fiduciary duty.[76]

Attempts to implement a more efficient prejudgment interest remedy can run afoul of practical and theoretical problems not taken into account in the assumptions of economic models. A case in point is furnished

---

71. In other words, a rational contract promisee would discount the value of a promise by a higher assumed interest rate on potentially withheld damages than a lender would charge the promisor. The promisor would therefore not purchase the right to withhold damages, i.e., would not take a loan, from the promisee.

72. A. Corbin, *supra* note 4, §§ 995, 1046; Restatement of Contracts § 337 (1932); Restatement (Second) of Contracts § 354. *See* Archer-Daniels-Midland Co. v. Paull, 313 F.2d 612 (8th Cir. 1963); Columbian Mut. Life Assur. Soc. v. Whitehead, 193 Ark. 598, 101 S.W.2d 455 (1937).

73. *See generally* Note, *Prejudgment Interest: Survey and Suggestion*, 77 Nw. U.L. Rev 192 (1982).

74. The Colorado statute *infra*, note 77, is an example.

75. D. Dobbs, *supra* note 61, at 169. (contrasting the award of interest as damages intended as compensation for the losses of the plaintiff with award of interest as restitution, intended to prevent unjust enrichment of the defendant).

76. *See* Friedmann, *supra* note 27.

by the state of Colorado, which has by statute[77] adopted the disgorgement remedy to measure prejudgment interest. The statute permits a promisee to recover from the promisor any gain that the promisor realizes from the wrongful withholding of money or property until it is paid or judgment is entered.

The ostensible purpose of the statute is to implement the disgorgement method of calculating prejudgment interest. But this principle has undergone a mutation in breach of contract cases. Courts applying the rule have forced defendants to disgorge the benefits that they have realized by reallocating performance resources rather than the benefits they have realized by withholding the damages payments. Given the propensity for breachers to withhold damages and incur liability for prejudgment interest, the new statute is tantamount to a general disgorgement remedy for breach of contract.[78] The story of how a simple prejudgment interest statute effected such a change in contract remedy law is a cautionary tale of the perils of legal conceptualism in the new age of law and economics.

The statute had its genesis in *Davis Cattle Co. v. Great Western Sugar Co.*,[79] an innovative application of disgorgement as a remedy for breach of a sale of goods contract. The Great Western Sugar Company was sued for breach of contract in a class action by its suppliers, some 3,600 growers of sugar beets. The growers had sold sugar beets to the company under a contract that fixed the price as a percentage of the company's net receipts from the sale of the sugar made from the beets. Although that price could not be determined until the beets had been processed and the sugar sold, the contract required the company to make an initial payment shortly after delivery of the beets. The initial payment was to be followed by interim payments over the course of the next eleven months, during which time the beets were processed and the sugar was sold.[80]

The contract required the initial payment to be at a rate that "shall be at the highest rate per ton [of beets] that the Company may deem to be justified taking into consideration anticipated returns from the sale of sugar and the sugar content of beets."[81] The company had

---

77. COLO. REV. STAT. § 5-12-102 (Supp. 1982). This statute is described and discussed more fully below.

78. *See infra* notes 79-93 and accompanying text. Colorado is apparently unique in this respect. Farnsworth, *supra* note 19, at 1369 ("In no jurisdiction do courts generally apply the disgorgement principle."); Friedmann, *supra* note 27.

79. 393 F. Supp. 1165 (D. Colo. 1975), *aff'd* 544 F.2d 436 (10th Cir. 1976), *cert. denied*, 429 U.S. 1094 (1977).

80. *Davis Cattle*, 393 F. Supp. at 1167-70.

81. *Id.* at 1169.

historically estimated what it would ultimately pay the growers and had paid about 85% of the anticipated total payment in the initial payment.[82]

In 1974 the sugar market was affected by several events that caused the price of sugar to skyrocket. The company paid the growers an initial payment much lower in relation to the (higher) price of sugar then its historical practice had been in more stable markets. The growers sued on grounds that the initial payment was so low as to breach the contract.[83]

The court found that, in making this payment the company did not consider the anticipated returns from the sale of sugar as the contract required, but instead selected a lower payment level in order to save interest costs it would otherwise have incurred in borrowing the money to make a larger payment.[84] The court held that the company breached the contract by fixing the payment in this way.[85] In an alternative finding, the court held that the company breached the contract by setting the initial payment in bad faith.[86]

Because full payment was made by the end of the processing year, the growers' damage from the breach arose from the delay in payment.[87] An interest award premised purely on the doctrine of expectation damages would have equalled the growers' economic injury resulting from the delay in receiving the initial underpayment. In Colorado, interest at the statutory rate supplanted expectation damages from breach of contract to pay money.[88] Under Colorado statutory law, however, the growers could recover no prejudgment interest because their claim for underpayment was unliquidated.[89]

The court met this obstacle creatively. Although agreeing that statutory interest could not be awarded on an unliquidated claim, the court held that under the common law, "moratory interest," or interest as damages, could be awarded on an unliquidated claim.[90] The court also

---

82. *Id.*

83. *Id.* at 1170.

84. *Id.* at 1178. The company argued unsuccessfully that the lower payment was essential to avoid the risk of overpayment, which might have occurred if the price of sugar dropped during the processing/sale period (as in fact it did).

85. *Id.* at 1175.

86. It did not help the company's argument on this point that, in attempting to prevent a hostile tender offer for its shares, the company had projected a net return on sugar much higher than the projection upon which the growers' initial payment was calculated. *Id.* at 1177.

Interestingly, the court did not allude to the good faith requirements of the Uniform Commercial Code, then in effect in Colorado and in at least some of the states in which the growers resided.

87. *Id.* at 1177-78.

88. Colo. Rev. Stat. § 5-12-102.

89. *Id.*

90. *Davis Cattle*, 393 F. Supp. at 1193-94. The trial court relied on an extensive

rejected the company's alternative argument that the growers were limited to the statutory rate of interest, then six percent. The court looked instead to restitution law, which measured damages by the amount of benefit to the wrongdoer resulting from breach.[91]

Even so, the path to judgment was not yet clear because the company was not shown to have made a profit on any money withheld. The court instead awarded as moratory interest the amount of interest expense the company saved by not borrowing against its bank lines of credit in order to make an initial payment at the required level.[92] The evidence indicated that the company would have had to pay at least 11.5% to borrow the extra money, and the court awarded the growers prejudgment interest at this rate.[93]

The *Davis* case appeared to establish a "pure" disgorgement remedy for breach of a contractual duty to pay money.[94] Breach of a contractual promise to pay money presents a special case in which the two duties, performance or payment of damages, are effectively, if not conceptually, identical. Thus, in *Davis* the same benefit accrued to the company from its withholding of performance and from its withholding of damages. Nor did the court have reason to clarify whether the "failure to pay money due" was the failure to perform or the failure to pay damages. Perhaps this ambiguity was responsible for the ensuing legislative failure to clarify the scope of the remedy in cases in which the withholding of performance and the withholding of damages diverged.

Soon afterwards, in 1979, the Colorado legislature adopted some aspects of the disgorgement remedy theory of *Davis* in an amendment

---

review of Colorado caselaw, including Bankers Trust Co. v. International Trust Co., 108 Colo. 15, 113 P.2d 656 (Colo. 1941), which held that interest as damages, and not as statutory interest, could be awarded for the tortious withholding of money or property. The plaintiff in *Banker's Trust* sought rescission of a purchase of securities induced by fraudulent misrepresentation. *Banker's Trust* was thus a garden variety restitution case, involving a tortious taking or detention of money. *Id.* at 33-34, 113 P.2d at 665. The *Davis Cattle* court applied this rule to a breach of contract case, holding that the measure of moratory interest is the gain or benefit realized by the wrongdoer. 393 F. Supp. at 1191.

91. Indeed, the language of the court in *Davis Cattle* suggests that the plaintiffs did not even have the option of recovering expectation damages, which is consistent with restitutionary precedent but not general contract principles. 393 F. Supp. at 1178.

92. *Id.* at 1194-95. Except where the wrongdoer saved an amount owed to the plaintiff, none of the cases cited by the court as precedent had measured the wrongdoer's benefit by the amount saved, only by the amount realized. The *Davis Cattle* court did not comment on this difference, which could be argued to be economically irrelevant.

93. *Id.* In fact, the evidence showed only that the company could have borrowed an additional $25 million at this rate. *Id.* at 1195. It was unclear whether the company could have borrowed sufficient money to make an adequate initial payment at any rate, since the company's available bank loan financing for the sugar beet payment was less than the initial payment amount that the court found to be required under the contract.

94. *Id. See supra* note 91.

to the Colorado prejudgment interest statute. The amended statute provides that when there is no agreement as to the rate of interest to be paid, creditors shall receive interest as follows:

> (a) When money or property has been wrongfully withheld, interest shall be an amount which fully recognizes the gain or benefit realized by the person withholding such money or property from the date of wrongful withholding to the date of payment or to the date judgment is entered, whichever first occurs; or, at the election of the claimant [interest on the amount withheld at the statutory rate of 8%] . . .
>
> (3) Interest shall be allowed . . . even if the amount is unliquidated at the time of wrongful withholding or at the time when due.[95]

The statute contains at least three striking features not present in the predecessor statute: prejudgment interest is permitted on unliquidated claims, the interest is measured by the wrongdoer's benefit, and the interest runs against a person withholding "money or property."

The statute suffers from a drafting ambiguity because of the uncertain meaning of "the gain or benefit realized by the person withholding such money or property." This can refer either to benefits resulting from the breach of contract or to benefits resulting from withholding payment of damages after breach, or perhaps to both.[96] *Davis* gives no guidance to interpretation because the benefits from the company's withholding of the initial payment and the benefits from the company's withholding of damages were the same. One is left with the question of what is the correct measure of gain or benefit where the breach involves failure to deliver something other than money.

This question was answered in *Great Western Sugar Co. v. Northern Natural Gas Co.*,[97] in which the same sugar company, now versed in the magic of moratory interest, essayed its hand as a plaintiff. The company purchased natural gas used in the processing of its sugar beets from a natural gas company. The gas was purchased under contracts that permitted the seller to interrupt service to the company whenever, "in its absolute discretion and without liability" to the buyer, the seller was "required to do so" in order to meet the demands of higher priority residential and commercial users.[98]

95. Colo. Rev. Stat. § 5-12-102.

96. *See infra* note 106.

97. 698 P.2d 769 (Colo. 1985), *aff'd on reh'g for damages sub nom.* Great Western Sugar Co. v. KN Energy, Inc., No. 87CA1477, slip op. (Colo. App. May 4, 1989) *aff'd* 778 P.2d 272 (Colo. App. 1989).

98. *Great Western*, 698 P.2d at 773.

During the winters of 1973 through 1979, the seller interrupted service to the company several times under the authority of the interruption provision.[99] During these interruptions, the company was required to use fuel oil, which was more expensive. To recover this extra expense, the company sued the seller, arguing that the interruptions breached the contract because they were not "required." The court found that the interruptions breached the contract and awarded the company damages of $3.6 million, representing its increased costs of cover in the fuel oil market.[100]

Relying on the new statute, the company also sought disgorgement of the benefit to the seller resulting from the breach.[101] The company offered two theories supporting such a claim, one based on the benefit to the seller from its retention of the gas withheld and one based on the benefit to the seller from its retention of the damages payments found to have been owing to the company.[102]

In support of the first theory, referred to as the "gas husbanding" model, the company introduced evidence that the seller had benefited from its withholding of the gas in several ways: the overall price of gas had increased, thus increasing the value of the withheld gas in the seller's hands; the gas withheld was available for sale at higher prices to residential users; and the seller had used the value of withheld gas to increase its rate base, and thus its income. From these increases in value, the model deducted the interest that would have been earned by the defendant on the contract price of the gas had it been sold to the buyer.[103] As of the time of the trial on interest in 1987, the court found the net benefit to be nearly $12 million.[104]

In support of the second theory, the company submitted three calculations of the seller's benefit on the damages withheld using, alternatively, the seller's internal rate of return on equity, the interest rates

99. The seller sought unsuccessfully to justify these interruptions by the national policy of conservation resulting from the gas crisis of of the 1970's. *Id.* at 778.

100. Great Western Sugar Co. v. KN Energy, Inc., No. 87CA1477, slip op. at 1.

101. *Id.*

102. 778 P.2d at 273; Slip opinion, *supra* note 97, at 1-2.

103. 778 P.2d at 274. The rate used to calculate this interest component was the average internal rate of return on investment achieved by the seller over the interest period. *Id.* While this measure is reasonable, reflecting a possible investment of the sale proceeds, it is not necessarily the "correct" rate. What, for example, if the seller had invested the sale proceeds by purchasing more natural gas? If the seller was indeed husbanding natural gas in anticipation of future price increases, such a purchase would seem natural. Presumably, the seller could have purchased at least as much gas as it sold to the buyer, and so would have realized benefits on the purchased gas equal to those that it realized on the withheld gas. This theory would result in no *net* benefit from the withholding.

104. *Id.* at 273.

that the seller paid for capital, and the rate of interest paid to the seller by its subsidiaries for money loaned to them. These calculations produced benefit figures ranging from $7.8 million to $8.7 million.[105]

The trial court held that the first theory must be applied and that damages must be measured by the benefit the seller gained by withholding the gas rather than by withholding damages. The court relied upon the statute's reference to the "gain or benefit realized by the person withholding such money or *property*."[106] From the $11.9 million in benefit accruing from the withheld gas, the court deducted the $3.6 million in compensatory damages and awarded the company $8.3 million in prejudgment interest.[107]

By selecting the gas husbanding model, the *Northern* court completed the transformation of the prejudgment interest statute into a general rule of disgorgement of the benefits to be gained from reallocating the performance resources. *Northern*, like *Davis*, does not even appear to give the plaintiff the option to recover the benefits the defendant realizes

---

105. Slip opinion, *supra* note 97, at 2.

106. COLO. REV. STAT. § 5-12-102(a). The statute is ambiguous: Does it mean "realized by the person [as the result of] withholding such money or property," or does it mean "realized by the person [who is] withholding such money or property?" The *Northern* court interpreted the statute in the first sense. The legislative history suggests that lawmakers considered the application of the statute to compel the disgorgement of benefits resulting from property withheld. *See* Transcript of Colorado Senate Judiciary Committee Hearingson S. 463, March 12, 1979, at 12 (analogy of convertor of a cow being liable for the value of milk obtained during its wrongful detention). When the phrase is read to apply to the "withholding" of property due under a contract, breach of contract is conceptually transmuted into the tort of conversion of the subject matter of the contract.

107. *Great Western*, No. 87CA1477, slip op. at 2. The trial court deducted the actual damages suffered by the plaintiff from the moratory interest award on the theory that "it is proper to deduct from the gross amount of such gain or benefit derived from the wrongful withholding of gas the amount [KN Energy] now has to pay in order to exercise its retention rights in that gas. Only in this manner can the Court fully recognize [KN Energy's] gain or benefit. A contrary finding would be punitive in nature as [KN Energy] has already paid for the gas that was not delivered to [Great Western]." Slip op. at 6. This reasoning was upheld on appeal. *Great Western*, 778 P.2d at 275-77.

An example may clarify the point. Consider a goods contract breached by the seller by sale to a third-party. Assume that the contract price is 100, the market price at the time of breach is 125, and the price realized by the seller from the sale to a third party is 150. The buyer, who covers at the market price, will recover expectation damages of 25, equal to the benefit of his bargain. If he also recovers as restitution the entire 50 that the seller realized on breaching the contract, the buyer will have a windfall of 25, receiving the equivalent of performance for 75.

The court in the *Northern Gas* case avoided this result by deducting the actual damages (the first 25 in the example) from the interest damages, which would lead to a total recovery of 50 in the example given. The court reasoned that the seller did not realize the full 50 as a result of the breach because the seller's "cost" of the breach was the 25 the seller had to pay the buyer as expectation damages.

from the withheld damages payment. Under Colorado's rule a defendant's liability for delayed payment is now wholly unrelated to the gain she realizes from delaying payment and the plaintiff's award is wholly unrelated to his losses from delayed payment. Thus, Colorado has achieved neither a disgorgement nor a compensation scheme of prejudgment interest. What it has done is to make efficient breach perilous if the breacher cannot accurately estimate the amount of her promisee's damages so that they can be paid at the time of breach.[108] While it creates a deterrent to the withholding of contract damages, the irrelevance of the disgorgement amount to the actual benefit of withholding will make the remedy either excessive or inadequate in most cases.

The Colorado courts have liberally construed the disgorgement remedy. It has been held to apply not only to contracts to pay money or deliver goods but to the "withholding" of any contractually required performance.[109] The courts construe withholding to commence whenever a right to damages arises.[110] Despite the statute's use of the word

---

108. There is no precedent yet on the effects of an inadequate but good faith payment of damages at the time of breach. Presumably, prejudgment interest would be due on the unpaid amount. But if, as *Northern* mandates, interest is computed on the withheld *property* rather than the withheld damages, defendant's attempted payment of damages will be unavailing and will not reduce the defendant's liability for prejudgment interest.

109. Benham v. Digby Trucking Co., 685 P.2d 249 (Colo. App. 1984) (insured's breach of contractual obligation to pay assessment to receiver of interinsurance exchange); Great Western Sugar Co. v. Kansas-Nebraska Natural Gas Pipeline Co., 661 P.2d 684 (Colo. App. 1982) (breach of contractual obligation to sell natural gas); Hott v. Tillotson-Lewis Constr. Co., 682 P.2d 1220, 1222-23 (Colo. App. 1983) (breach of contract to construct a house).

In Isbill Assoc. v. City and County of Denver, 666 P.2d 1117 (Colo. App. 1983), applying the statute to the damages due on a lessee's tort claim against its lessor for water damage, the court noted that the legislative history of the statute expressed an intention to benefit all plaintiffs, not just those seeking recovery of money due on a contract to pay money. *Id.* at 1122.

The federal district courts sitting in Colorado have refused to apply the statute as broadly as has the state appellate court. Lowell Staats Min. Co. v. Pioneer Caravan, Inc., 645 F. Supp. 254 (D. Colo. 1986) (refusing to apply the statute to "garden variety breach of contract cases."). *Id.* at 259. *See also* EEOC v. Trailways, 28 FEP 552 (D. Colo. 1981) in which Judge Winner, the author of the opinion in *Davis*, refused to apply the statute to an employee's claim for wages in a wrongful discharge case, stating: "To extend to a wrongful discharge case the concept of awarding interest for money wrongfully withheld would mean that every time anyone won a suit for breach of contract, monetary interest would have to be awarded. Having spent the time I did in studying prejudgment interest before writing *Davis Cattle*, I am unable to conclude that prejudgment interest should be awarded on the unliquidated amount here involved, and I say this with full awareness of the amendment to the Colorado interest statute. [citation omitted]." 28 FEP at 553-54. Unfortunately the judge did not explain how his opinion in *Davis* should be limited so as not to apply to all suits for breach of contract.

110. *See Isbill Assoc.*, 666 P.2d 1117. (applying the statute to a tenant's tort claim for property damage against its landlord).

"wrongfully," plaintiffs have not been required to show that the withholding was tortious or in bad faith.[111] As a result, the statute applies to any breach of contract.

V. THE PROBLEM OF CAUSATION: LEGAL CONCEPTUALISM AND ECONOMIC ANALYSIS

This part analyzes the formidable transaction costs of calculating the benefits to be awarded under a disgorgement remedy. The foregoing analysis has proceeded on the assumption that the "benefits of breach" or the "benefits of withholding damages" could be identified and measured. These phrases imply a direct causal relationship between the breach or withholding and some benefit. As Professor Farnsworth has pointed out, the causal relationship between breach and benefits is often simply assumed in restitutionary theory.[112] Likewise, the misnomer "efficient breach" implies a false causal relationship. The benefits related to efficient breach are not "caused" by the breach. The benefits result from some allocation of performance resources, an allocation that itself may or may not also "cause" the breach. These relationships do not coincide with traditional legal conceptions of causation, which erect intractable problems in the calculation of disgorgement benefits.

The relationship among the breacher's allocation of resources, the breach, and some benefit may be quite tenuous. Both before and after contracting, a promisor may make allocational decisions that make performance less likely or impossible. Promisors may accept additional contracts, invest potential performance resources in other projects, and take various other risks that jeopardize the performance of the contract. These decisions presumably benefit the promisor.

A disgorgement remedy aimed at these benefits would have to satisfy some standard of legal causation.[113] Classically, this would require showing that the benefits were caused by the contract breach, for it is the breach and not the resource allocation upon which a plaintiff's claim is based. When the benefits precede the breach, this burden cannot be met.[114] The Colorado statute, by employing an explicit temporal condition,

---

111. *Benham*, 685 P.2d 249 (breach of contractual obligation to pay assessment to receiver of interinsurance exchange).

112. Farnsworth, *supra* note 4, at 1343.

113. On the meaning of legal causation see H. Hart & A. Honore, Causation in the Law (2d ed. 1985). On the role of cause in economic analysis *see* Coase, *supra* note 21; Landes and Posner, *Causation In Tort Law: An Economic Approach*, 18 J. Legal Stud. 325 (1989); Culp, *Causation, Economists, and the Dinosaur: A Response to Professor Dray*, 49 Law & Contemp. Probs. 23 (1986).

114. For example, in performing Contract A a promisor may suffer damage to equipment, making it impossible to perform Contract B. One might analogize this to a sale of goods to the Alternative Purchaser, so that the Promisor's profit from Contract A is a "benefit" obtained "as a result of" the breach of contract B.

precludes recovery of benefits that do not arise "from the date of wrongful withholding" until payment or judgment.[115] If the date of withholding is the date of performance, and if the wrongdoer has fully realized the benefit *before* the date of performance, disgorgement of the benefit is presumably unavailable.[116] The need to tie the disgorgement remedy causally to the breach rather than to the decision to reallocate (and breach) creates the opportunity for strategical evasion of the pre-judgment interest remedy. If the breacher can time her transactions so as to realize the benefit of withholding before the breach, she can avoid disgorgement liability for the withholding, although the practical effect may be identical to a realization of benefit after the withholding. Economically, the results are identical but the requirements of legal causation lead to disparate treatment.

A. *Calculation of the Benefits from Reallocating Performance Resources*

1. *Identified Performance Resources.* At times, such as in the sale of goods existing at the time of the contract, the parties may informally recognize an intention to use specific performance resources in the performance of the contract, yet the buyer may never acquire a right to recover for their conversion.[117] When these goods are reallocated to some other use, calculation of the benefits of the reallocation would seem simple.

But this is not always the case. Consider the well-known case of *Acme Mills & Elevator Co. v. Johnson.*[118] Defendant Johnson, a farmer, contracted to sell 2,000 bushels of wheat to the plaintiff Acme, a grain dealer, at $1.03 per bushel to be delivered at harvest time. The market rose sharply, and the alert Johnson sold his wheat at $1.16 to a different buyer. By the time of harvesting, however, the market had dropped to $1.00 per bushel. When Johnson failed to deliver, Acme sued for the extra $.16 per bushel Johnson realized on the second sale.[119] This remedy

---

115. Colo. Rev. Stat. § 5-12-102(a).

116. Actions that amount to repudiation might be construed as "wrongful withholding," however, as they constitute breach giving the promisee the immediate right to damages. U.C.C. § 2-610.

117. The concept of identification to the contract is used in the Uniform Commercial Code, § 2-501, as the process whereby a buyer of goods obtains a special property and insurable interest in the goods. Identification occurs for existing goods at the time of contracting (U.C.C. § 2-501(1)(a)); and for future goods when the seller ships them or marks them as the goods to which the contract refers. U.C.C. § 2-501(1)(b).

118. 141 Ky. 718, 133 S.W. 784 (1911).

119. *Id.* Under the Uniform Commercial Code, the *Acme* breach would not have been efficient if Johnson had informed Acme at the time of his sale to the alternative

was refused, the court measuring damages as the difference between the market and contract price at the time and place for delivery. Under this measure Acme had suffered no loss (and had in fact benefited from Johnson's failure to perform).

Had a disgorgement remedy been available, the benefit of the reallocation would seem to have been thirteen cents per bushel, the price the breacher received for the goods less the contract price. Yet, Johnson could have purchased grain in the market at $1.00 to satisfy his obligation to the buyer. If he had, he could have realized his profit and fulfilled his contract as well (earning three more cents a bushel in the process.) One could argue that the benefit caused by the breach was, therefore, only the amount by which this market price exceeded the contract price: in the *Acme* case, zero.

Professor Farnsworth's leading commentary on the application of the disgorgement remedy to breach of contract would apply such a version of legal causation in determining the benefit resulting from breach.[120] Professor Farnsworth calls for a "strict cause in fact" requirement that would limit disgorgement to benefits as to which breach was essential in a "but-for" sense, i.e., benefits that the breacher could not have obtained without the breach. Benefits would be measured as the least amount of savings that the breacher could have realized by not performing the contract. Thus, if instead of breaching the contract, the breaching party could have purchased goods in the market in order to satisfy the alternative transaction, disgorgement would be limited to the amount the breacher saved by not doing so.[121]

This approach would have led to a different result in *Northern*. The gas seller could have both realized its gain and fully performed at a minimal cost. At the time of breach, for example, it could have purchased additional gas for sale to the buyer or for storage to replace the gas it sold. Its savings would have been the purchase price of gas at the times of breach less its sales price from the sale to the plaintiff. The breach was in all likelihood not necessary to the gain the seller realized, and Farnsworth's method would have precluded a disgorgement award.

---

purchaser that he would not be able to perform his contract. Under Section 2-713, Acme's damages for Johnson's repudiation would be the difference between the market price at the time Acme learned of the breach, presumably $1.16, and the contract price. These damages would have eliminated Johnson's profit.

120. *See* Farnsworth, *supra* note 4.

121. *See id.*, at 1344 (possibility of seller's market purchase of substitute widget to sell to alternative buyer destroys causal link between breach and gain). Ironically, by permitting disgorgement only where the breach was necessary to the gain, Farnsworth's strict cause requirement absolves promisors who unnecessarily breach contracts and punishes those whose breach was necessary to the gain.

A corollary of Farnsworth's strict cause requirement is the doctrine of joint cause. Any part of the breaching party's gain that results from her own skill or efforts must be deducted from the total gain in order to calculate the amount of gain "caused" by the breach. Farnsworth analogizes this principle to the deduction of brokerage or similar costs or fees from a wrongdoer's gross profit in cases of fiduciary disgorgement.[122]

This doctrine, too, would have altered the *Northern* result. The $11.3 million gain in value of equity that the seller realized was due in large part to the seller's contribution to the value of the gas in storing it, in using the gas in rate-making, and in diverting the gas to higher paying customers. In addition, the seller took market risks of price changes in gas. Deduction of the value of the seller's contribution would have greatly reduced the award.

To allow breachers the benefits of Farnsworth's imaginary or hindsight efficiencies will not encourage them to act efficiently at the time of breach. But such an approach is preferable to the *post hoc ergo propter hoc* fallacy exemplified in *Northern*. In neither case are the benefits of reallocation causally linked to breach with sufficient vigor.

2. *Non-Identified Performance Resources*. To attempt to trace the benefits from the allocation of resources that have not been designated to the contract is even more of an exercise in fiction. For example, in *Northern*, the court purported to value hypothetical gas that was not sold to the buyer by deeming it to have been stored for years in an underground storage facility, where it generated an increase in the seller's equity. But the seller could plausibly have argued that the "buyer's gas" was never purchased by the seller, or that it was sold to residential customers, or even that it was sold in later years to the buyer itself. All these stories would have been equally "correct" in the absence of any identification of specific gas to the contract. In selecting from among the fictional futures of the fictional gas, the court has no economic or other principle to guide it. A remedy so unmoored to reality can achieve neither efficiency nor justice.

3. *Measuring Benefit by "Savings" Resulting from Breach*. An even deeper level of fiction is reached when courts consider, as did the court in *Davis*, that a breacher benefits from breach by the amount of "savings" realized from failure to perform. When performance resources have never been allocated to the contract, one cannot speak meaningfully of their "re"allocation to higher valued uses. For example, consider a manufacturer who sells goods that may contain defects that will cause predictable losses. The manufacturer should spend resources on quality

122. *Id.* at 1347-48.

control to the point of diminishing return, measuring amount spent against loss avoided. At this level of precaution the manufacturer is acting efficiently and the joint cost of the product is minimized even though some defects will occur.[123]

If the inevitable defective product breaches the contract, how is the "benefit of the breach" to be calculated? The manufacturer did not realize a gain from any obvious intentional breach, but a court applying the "savings" rationale might ask what the promisor would have had to spend in order not to breach, as did the court in *Davis*. The amount the manufacturer might have saved by not spending enough to reduce defects to zero percent might approach infinity. Disgorgement of this amount would give the plaintiff a property interest in the hypothetical dollars that the manufacturer did not spend, nor even considered spending. The principle would also reach, for example, breaches in which the promisor acted efficiently by overbooking or overscheduling production. The "benefits of breach" would equal the cost of the excess capacity the promisor would have had to maintain to assure performance.

A general disgorgement remedy would not be efficient in such cases because it would lead the promisor to overinvest in precautionary care and raise the joint costs of the contract. Promisees profit when the joint costs are minimized and would be charged more for a general disgorgement remedy than it would be worth to them.

4. *Informational Transaction Costs of Calculating the Breacher's Benefit.* The administrative and uncertainty transaction costs of calculating the benefit under a disgorgement remedy will usually exceed those of calculating a loss under an expectation damages remedy.[124] In the case of breach of contract, the injury is the shortfall between the breach situation and the promisee's expectation; the benefit of breach is the improvement of the promisor's condition over that which performance would have created. The benefit is often more costly to calculate than the injury because of the parties' unequal access to relevant information. The party in possession of the best information about the injury, the contract plaintiff, is the one with the burden of proving it. The party in possession of the best information about the benefit, the breaching promissor, is the one resisting the claim. The same inequality applies to information about the hypothetical world that would have ensued upon full performance. The promisee knows what he was expecting from performance, but not what the promisor was expecting. Finally, a prom-

---

123. Cooter, *supra* note 8.

124. It has been noted that courts tolerate more uncertainty in calculating losses than in calculating benefits for restitutionary purposes. Levmore, *Explaining Restitution*, 71 Va. L. Rev. 65, 72 (1985).

isor may subjectively value reallocation of performance resources in ways that the promisee can never prove.[125] Subjective benefits of breach would be immune to disgorgement, although they would be very real.

These information inequalities are magnified by the conjectural nature of the calculation of the breacher's gain. The transaction costs of such a calculation are lower for the promisor than for the promisee. This suggests that the promisor is the least cost avoider of the risk of her benefit from a reallocation of performance resources. Assigning the remedy, and therefore the risk, to the promisee is inefficient.

B. Calculation of the Benefits from Withholding Payment of Damages

Even where the disgorgement remedy is limited to the benefits from withholding the damages payment, legal causation does not furnish an acceptable measure. Unlike the typical fiduciary case, a breach of contract rarely involves an identifiable fund from which to trace benefits, and the resulting calculations are fictional and arbitrary.[126]

An example is furnished by the benefit measurements rejected by the court in *Northern*. The plaintiff introduced evidence of the seller's average investment return, its lending rate to its subsidiaries, and its average borrowing rate.[127] These produced measures of benefit that differed by almost $1 million. But which of these is the "correct" measure of the seller's benefit from withholding payment of damages?

One possibility requires the promisor to disgorge the benefit of the alternative transaction, the result reached in *Northern*. Such an approach recognizes that breach damages may be deemed "invested" in any of the promisor's property, including the profit-making activity that replaced performance of the contract. But the *Northern* court did not compare the amount of the unpaid damages to the amount the seller had invested in the withheld gas. The reallocation is only one of the plausible in-

---

125. This would not only include the rare case when the promisor takes malicious delight in breaching, but would include more common cases, as when the promisor consumes the performance resources herself, i.e., by going fishing instead of painting plaintiff's barn. A disgorgement of this benefit would require a measure of the promisor's subjective valuation of her leisure. While a plaintiff may seek to prove *his* subjective valuation of performance as a measure of expectation damages, it is not likely that he could succeed in proving his adversary's subjective state. *See also* Tredennick and Cairns, *Colorado's Prejudgment Interest Statute: Potential for Market Rate Interest* 12 Colo. Law. 1605 (1983).

126. On the problem of tracing, *see* G. Palmer, *supra* note 26 §§ 2.14, 2.16; Restatement (Second) of Trusts § 202; Restatement (Second) of the Law of Restitution, Tentative Draft No. 2 (April 6, 1984) §§ 34-41.

127. *See* Farnsworth, *supra* note 4 (referring to the cost of substitution, the least cost the wrongdoer would have incurred in performing the obligation).

vestments the promisor might have made with the damages payment.

A second method, suggested by Professor Farnsworth's cost of substitution analysis, gives the breacher the benefit of the doubt by measuring the benefit as the lowest cost that the breacher could have incurred in discharging the damages obligation.[128] This method assumes that if the breacher were forced to pay damages on time, she would have done so at the lease expense to herself, a not unreasonable assumption. The seller in *Northern* could have liquidated its least profitable investment or borrowed at the lowest available rate in order to pay the damages.

Yet this method involves the problem of retrospection. If the breacher's benefit is to be measured by her return on some investment, and if the investment chosen is to be the one with the least return, then as of when will that return be measured? The breacher can wait until trial, pick her least profitable investment, and perhaps pay nothing in interest. This method would encourage the breacher to invest in hedge-type speculations in which "losers" are expected to be balanced against "gainers." After entry of judgment, the plaintiff can be given the "benefit" of whichever investment dropped while the breacher will enjoy the winner. A possible solution is to attempt a judicial, *ex post* determination of the expected return of the investments at the time of breach, but this method creates obvious valuation problems and attendant costs.

A third approach is to measure the benefit to the breacher as if the breacher had invested the withheld damages in the highest return investment in her "portfolio." Such a measure will give the breacher an efficient incentive to pay damages upon breach and the breacher will determine the most efficient way to manage her portfolio so as to make that payment. Again the problem of retrospection will arise, permitting the plaintiff to "cherry pick" from the defendant's portfolio at the time of judgment. In this case, however, the defendant will probably avoid the risk by paying the damages on time.

Even if a fund from which the damages payment would have been made could be identified with certainty, a court attempting to apply a disgorgement remedy would still face the problem of tracing. Professor Farnsworth argues that a court applying a disgorgement remedy should refuse to "trace" ill-gotten gain through transactions that enhance its value.[129] While it might simplify judicial determinations, this approach would certainly stimulate strategic behavior in potential disgorgement defendants, who would simply "launder" money before investing it. If disgorgement is an efficient remedy, courts must engage in tracing in order to make the remedy work.

---

128. *Id.*
129. *Id.*

VI. CONCLUSION

The disgorgement remedy responds to our sense that no one should be able to profit from her own wrong. But things change when wrongs are very profitable. As a general remedy for breach of contract, disgorgement would violate the principle of cost avoidance in all but a few specific situations. Whether or not the storied costs of post-contract renegotiation would exceed the gains in preventing inefficient breach, any such excess would soon be dwarfed by the transaction costs of judicial calculation of benefits from breach. Our ideas of legal causation are poorly suited to this calculation.

As a remedy for withholding damages, disgorgement is more promising, although problems of calculating benefits remain. A judicial rule of thumb that selected the highest rate of investment or borrowing in the breacher's portfolio would seem appropriate. Approaches, such as Colorado's, that award the benefits of reallocation as a remedy for withholding damages are better left to the place where unsuccessful legislative experiments go.

Defending Purchase Money Security Interests Under Article 9 of the UCC From Professor Buckley

PAUL M. SHUPACK*

I. INTRODUCTION

At least since Karl Marx, economists have moved from theory to practice with what, in hindsight, appears to have been unseemly haste. [1] Those who analyze law from the view of economics have tended to share this trait. Professor Francis H. Buckley, in his article, *The Bankruptcy Priority Puzzle*,[2] recommends immediate changes in our legal practice in light of the insights resulting from his theoretical inquiry. This Article argues that no persuasive reasons exist to accept the changes Professor Buckley proposes for Article 9 of the Uniform Commercial Code ("UCC"). His theory does not translate into practice as easily as he appears to believe it does.

Professor Buckley's argument, developed in counterpoint to a literature which asks why the law permits any secured lending,[3] questions any statutory regulation of secured debt. He observes that "management is the decisionmaking body best able to determine the particular firm's value-maximizing secured debt policy."[4] He then asks why the UCC imposes what he calls mandatory barriers to management discretion in creating secured transactions. While he defends some of these mandatory rules, he finds inadequate the arguments conventionally used to justify

---

* Professor of Law, Cardozo School of Law, Yeshiva University; A.B. Columbia, 1961; J.D. University of Chicago, 1970. I wish to thank my colleagues William Bratton and David Carlson for their help with this article.

1. "The philosophers have only *interpreted* the world in various ways; the point, however, is to *change* it." K. MARX, THESES ON FEUERBACH (*reprinted in* 2 K. MARX & F. ENGELS, SELECTED WORKS at 405 (1962).

2. Buckley, *The Bankruptcy Priority Puzzle*, 72 VA. L. REV. 1393 (1986).

3. This literature includes Jackson & Kronman, *Secured Financing and Priorities Among Creditors*, 88 YALE L.J. 1143 (1979); Jackson & Schwartz, *Vacuum of Fact or Vacuous Theory: A Reply to Professor Kripke*, 133 U. PA. L. REV. 987 (1985); Kripke, *Law and Economics: Measuring the Economic Efficiency of Commercial Law in a Vacuum of Fact*, 133 U. PA. L. REV. 929 (1985); Levmore, *Monitors and Freeriders in Commercial and Corporate Settings*, 92 YALE L. J. 49 (1982); Schwartz, *Security Interests and Bankruptcy Priorities: A Review of Current Theories*, 10 J. LEGAL STUD. 1 (1981); Schwartz, *The Continuing Puzzle of Secured Debt*, 37 VAND. L. REV. 1051 (1984); Scott, *A Relational Theory of Secured Financing*, 86 COLUM. L. REV. 901 (1986); White, *Efficiency Justifications for Personal Property Security*, 37 VAND. L. REV. 473 (1984); Shupack, *Solving the Puzzle of Secured Transactions*, 41 RUTGERS L. REV. 1067 (1989).

4. Buckley, *supra* note 2, at 1469.

others. Among the rules he finds most open to question are those creating superpriority for purchase money security interests ("PMSI"). Under the UCC, a first-in-time[5] secured lender who wants to use the debtor's after-acquired property as further security for a loan cannot be certain that the debtor's after-acquired property will be effectively available as collateral. Despite the terms of any contract between the debtor and a first-in-time secured creditor, the debtor may acquire that property subject to a PMSI, to which the UCC awards a priority ahead of the first-in-time secured party. Professor Buckley argues that this inability of creditors and debtors to bargain away the consequences of the statutory scheme imposes unnecessary costs on first-in-time secured creditors and their debtors.

Against Professor Buckley's conclusion that the phenomenon of PMSI superpriority cannot be justified on efficiency grounds, this Article argues that the existing legal order can be rationally defended. In this type of Kaldor-Hicks efficiency discussion, the absence of data prevents the possibility of proof.[6] What follows is very much in the spirit of Professor Buckley's article. It offers reasons to believe that if data did exist, it would show that PMSIs would be efficient, even using Professor Buckley's special test for what he calls mandatory rules, though it also takes issue with the adequacy of that test to determine any question of public policy.

This Article shows that enough debtors and secured creditors with after-acquired property clauses in their loan agreements will want to allow for PMSI superpriority that the probable transaction costs saved by the existing arrangements will outweigh the probable costs to debtors and secured parties resulting from these restrictions. That demonstration alone makes Professor Buckley's argument somewhat implausible.

Moreover, Professor Buckley does not consider the costs that his alternative regime of free contract would impose on both first-in-time and second-in-time secured lenders. The first-in-time secured lenders would find that they had added drafting and collateral realization costs under a free contract regime compared to those costs under the UCC. Under the UCC, first-in-time secured lenders potentially bear costs of

---

5. By first-in-time, I mean the secured lender who first makes a claim in a security agreement to collateral that qualifies as after-acquired property. Until the debtor has rights in the collateral, no secured party can have a claim to the collateral, and, as a result, the issue that this Article addresses, if described precisely, is to which of two simultaneously-attaching security interests does the law give priority.

6. For a transaction to be efficient in a Kaldor-Hicks sense, it must be one in which the utility gains of the winners exceed the utility losses of the losers, but there is no need for the winners to compensate the losers. *See, e.g.*, A. FELDMAN, WELFARE ECONOMICS AND SOCIAL CHOICE THEORY, 142-44 (1980).

certain types of debtor dishonesty. Professor Buckley notes that his regime of free contract will eliminate these risks to first-in-time secured lenders, but he does not mention that these risks are simply transferred to second-in-time secured lenders. There is no saving here, and in fact there is some loss because second-in-time lenders are less able to protect themselves against these risks. Once this more complex accounting is done, Professor Buckley's argument for the relative efficiency of a free contract regime becomes increasingly implausible.

II. Professor Buckley's Test for "Mandatory Rules"

When Professor Buckley analyzes what he calls mandatory legal rules in the UCC, he introduces a novel test by which these rules may be judged. In his view, to justify these mandatory legal rules, it is not enough to show that the effect of this type of rule creates efficient behavior. One must also show that "transaction costs or other barriers would prevent the parties from devising optimal priority rules in non-mandatory regimes."[7]

Professor Buckley's example of a "mandatory" rule is that part of UCC Section 9-307(1) which states, "A buyer in the ordinary course of business . . . takes free of a security interest created by his seller even though the security interest is perfected and even though the buyer knows of its existence."[8] This subsection gives buyers in the ordinary course of business rights superior to those of the seller's secured lenders in any goods bought from the seller.

Professor Buckley is satisfied that this statutory provision is fully justified by his novel efficiency test. Under that test, it is not enough that the parties, if left to their own devices, want to reach a result equivalent to the one reached by UCC Section 9-307(1). An additional element must be satisfied. If Section 9-307(1) did not exist, the parties would find it difficult and expensive to reach its result. Without mandatory superpriority, the screening process that buyers would have to undertake, *e.g.*, the review of the seller's financing documents, "seems extremely inefficient. Mandatory superpriority rights provide a simple solution to the problem."[9]

It is common to observe that legal rules, and especially contract rules, have the effect of providing ready-made patterns for action. These ready-made patterns create a savings for those who use them, modest costs for those who contract out of them, and potentially major costs for those who should have contracted out of them, but failed to do

---

7. Buckley, *supra* note 2, at 1452.
8. U.C.C. § 9-307(1) (1972).
9. Buckley, *supra* note 2, at 1454.

so. The difference between "many" and "all" justifies preserving the possibility of freedom of contract even when one is certain that the ready-made pattern is one that most would prefer. Should the legislature *require* the pattern, the requirement would prevent those who did not want it from contracting out of it. Professor Buckley argues that before a legislature requires a rule which results in efficient behavior, it ought to satisfy itself that some barrier exists to prevent the contracting parties from reaching the result they are assumed to want.

There are two problems with this otherwise attractive test. First, to the extent that a legislature does not make a rule "mandatory," it may well end up imposing costs on contracting parties. To the extent these costs exist, they must be weighed against the costs imposed on those who dissent from the mandatory rule. Second, the idea of a mandatory rule is itself somewhat problematic, especially in light of the way Buckley uses his own term.

If the meaning of a term may be understood from its use, then we can understand Buckley's concept of a "mandatory" rule through an examination of Section 9-307(1), which he has declared to be a good mandatory rule. While Section 9-307(1) gives superpriority to a buyer in the ordinary course of business, any buyer who would qualify as a buyer in the ordinary course of business could, by explicit contract, waive his right to take goods free of the security interest created by his seller. If a generous buyer did so, courts would uphold the consequences of such a waiver for buyer, seller, and seller's secured party (absent of course the defenses of unconscionability and the like which would be suggested by such an altruistic act in the midst of an ordinary commercial transaction). Calling this type of rule mandatory uses that word in a sense that differs from its ordinary usage.

When Professor Buckley uses the term "mandatory," he would have us look not to the relationship of buyer and seller, but rather to the seller and his secured lender. He would have us ask what the effect is under the law if a seller, whose business made his buyers "buyers in the ordinary course of business," were to contract with his secured party to seek a waiver of the effect of UCC Section 9-307(1) on each sales transaction. Again, nothing in the law would prohibit the parties from doing so or from ordering their affairs in accordance with this type of contractual system. The seller may find few buyers willing to purchase goods agreeing to this type of waiver, but if the seller can find them and the buyers consent, then the law is satisfied.

Problems arise only when the seller breaches his contract with the secured lender and sells goods without obtaining the waiver called for under the contract with the secured lender. In this event, Section 9-307(1) will deprive the secured party of any remedy against the goods. At this point, however, the secured creditor still would have a claim

against the seller based on breach of contract, a legal remedy which concededly she neither wanted nor contracted for.

To understand what Professor Buckley means by a mandatory legal rule, one must ask whether the so-called mandatory legal rule has the effect of making contractual provisions unreliable in their effect in the event one of the parties breaches the contract. The rule creating mandatory superpriority for buyers in the ordinary course of business makes unreliable any secured party's attempt to provide by contract for security interests that survive sale by her debtor of the goods serving as collateral. The law does not forbid parties from so contracting. Nor does the law leave the creditor suffering from breach of such a contract without remedy for that breach. The law does, however, leave the aggrieved party without the contracted-for remedy, thus imposing costs that the aggrieved party will find unacceptable. There is thus no difference in kind between an ordinary background rule of law, filling the gaps in the absence of agreement of the parties, and a rule of law which is "mandatory." A "mandatory" rule, in Buckley's usage is simply one that creates a result which imposes unacceptable costs on those who dissent from the rule.

III. Professor Buckley and the Purchase Money Security Interest

Using his test, Professor Buckley questions those provisions of Article 9 of the UCC law that create mandatory superpriority for PMSIs.[10] Under the UCC, a lender who takes a security interest in the debtor's after-acquired collateral cannot prevent the debtor from creating a secured lender second-in-time, whose interest in the after-acquired collateral will be prior in right to that of the secured creditor who is first-in-time. By having PMSI status, a second secured lender can effectively remove the benefit of the first secured party's after-acquired property clause.

In Professor Buckley's view, "[b]ecause security interests in after-acquired property give rise to readily apparent efficiencies, the exception for PMSI superpriority appears to require a countervailing allocational justification."[11] Professor Buckley then examines the "four basic theories" defending PMSI superpriority, and finds each wanting. A secured creditor with an after-acquired property clause has a situational monopoly, but, says Professor Buckley, if the law did not create PMSI superpriority, "any inefficiencies associated with an issue of after-acquired property interests [could be] cured *ex ante* in the first loan, which could itself provide for PMSI superpriority if it were efficient to do

10. *Id.* at 1462.
11. *Id.*

so."[12] The second justification for PMSI superpriority, which, as Buckley observes, is a variant of the first argument, notes that if a secured creditor with an effective after-acquired property clause is certain that the debtor's existing line of business will generate revenues sufficient to pay off the loan, then the secured creditor will have disincentives to permit the debtor to enter into high risk, high return new ventures. Buckley does not agree that the situation generates these sorts of disincentives because, as he notes, the first creditor can always make a second loan at an interest rate different from that of the first loan.[13] To the extent that first-to-file creditors do take advantage of their situational monopoly and engage in strategic behavior, *ex ante* contracting by the debtor reserving the right to create PMSI superpriority again operates as an effective prevention of the evil the statutory PMSI superpriority is supposed to cure.[14]

The third and fourth arguments that support statutory PMSI superpriority presuppose secured creditors with significant differences among themselves concerning their information about the debtor or the collateral. The logic of this theory leads to the conclusion that an after-acquired property clause benefitting such a monitoring secured creditor leads to a "sapping" of "the most plausible monitor's incentive to police the debtor's behavior,"[15] which in turn can be controlled by PMSIs. Buckley observes that if the problem is the loss of incentive to monitor by the original secured party, that incentive problem can be solved directly by recreating pre-UCC rules which conditioned the effectiveness of after-acquired property clauses on the creditor's monitoring of the after-acquired property.[16]

The other branch of justifications based on the creditor information differentials also contains the explanation that Professor Buckley finds most plausible as to why the law might favor PMSIs. To the extent that creditor monitoring of particular items of collateral guards against debtor misbehavior, and to the extent that creditor capacity to monitor is a function of the precise collateral involved, then the PMSI, with its capacity to tie a specific creditor to specific collateral, aids in bringing about that assignment at low cost. (Similarly, to the extent that creditors

---

12. *Id.* at 1463.

13. *Id.* at 1464.

14. *Id.* at 1464 n.158.

15. *Id.* at 1464-65.

16. *Id.* at 1465. Buckley's account uncovers yet another obscurity in the monitoring theory of secured transactions. If the justification for security interests is at bottom the benefit the monitor performs for herself and the general creditors, it is not at all obvious how the general creditors are helped by monitors whose advantage lies in the collateral rather than in the debtor.

have differing knowledge of the value of assets on default, it is likely that sellers of the collateral will also be in a better position to value the collateral in the event of default) The PMSI again operates to bring about an easy assignment of the "right" collateral to the "right" creditor. To both of these points, Buckley gives his familiar answer—to the extent that such regimes do exist, then if allowed to do so by law, debtors would agree *ex ante* with their first creditors to retain the right to create these sorts of PMSIs.[17]

IV. Defending PMSI Superpriority from Professor Buckley

With respect to three of the four objections he notes, Professor Buckley observes that if the parties wanted PMSI superpriority, and if the law did not already give it to them, they could create it by contract. The repetition in his analysis of the phrase "if the debtor wants PMSI superpriority, he could contract for it *ex ante*" should stand as warning that the analysis cannot be complete. So long as the analysis concerns reality, where contracting is itself a costly affair, the recognition that parties will repeatedly contract *ex ante* to bring about a particular result must suggest the possibility that a legal rule which brings about exactly that result is one that will not only be efficient on Kaldor-Hicks principles, but will also reveal a problem with Professor Buckley's additional test.

If enough people want the result required by a rule, then the commonplace Kaldor-Hicks efficiency account for having a rule at all

---

17. *Id.* at 1465-66. In his discussion of the possible advantages of the UCC's system of PMSI superpriority, Professor Buckley notes that creditor advantage with respect to collateral could create efficiencies by decreasing screening costs for all creditors. He then goes on to say, in an uncharacteristically opaque passage:

> Moreover, it is not at all clear that PMSI superpriority decreases screening costs. PMSI superpriority may compound uncertainties in the valuation of the debtor's assets. It requires the debtor and the after-acquired property financer to estimate not merely the anticipated value of the firm on default but also the risk that anticipated collateral may be lost to subsequent PMSI lenders. In this way PMSI superpriority may actually increase screening costs and thus the cost of credit.

Id. at 1466.

The passage fails to tell the reader against what measuring point there is increase and decrease. Presumably the increases and decreases are to be measured against a regime of free contract. Even if it were true that screening cost savings created by mandatory PMSI superpriority were outweighed by the screening costs, it does not necessarily follow that the cost of credit would "actually increase." This passage concludes a discussion in which Professor Buckley has described three other reasons to believe that mandatory PMSI superpriority might decrease the cost of credit. He has said that as to two of them, they plausibly suggest reasons to believe that PMSI superpriority does in fact reduce the price of credit. Thus even if screening costs were increased by mandatory superpriority, the reductions in costs of credit described in the other two theories might well mean that mandatory PMSI superpriority nonetheless does reduce the price of credit.

comes into play. The saved transactions costs for those desiring the result will outweigh the costs of those who now need to contract out of the result. If, however, the rule is a mandatory rule, the costs of contracting out of the rule are presumed to be so great that, as a practical matter, everyone is stuck with the rule. Instead of having the inconvenience and cost of contracting out of a rule's effect, those not wanting the rule now have the full cost of that unwanted effect. It is this additional cost that must be weighed against the savings to those contracting parties who, in the absence of mandatory rule, would have to provide *ex ante* for its result. Even so, there remains the possibility that the convenience of the many will, at some level, provide a Kaldor-Hicks justification for a mandatory rule. This result would occur if the costs the many would incur to contract into the rule's result would exceed the costs to the dissenters of having to follow the rule. It is precisely this calculation that Professor Buckley's special test does not permit. Yet there does not appear to be any reason, other than an *a priori* preference for freedom of contract, no matter what its costs to members of society, for excluding this type of calculation from a Kaldor-Hicks efficiency analysis.

In the absence of data, weighing these two sets of values against each other is admittedly difficult. If, however, it can be shown that hardly any contracting party is likely to want a result other than that called for by the mandatory rule, and if it can be shown in addition that those few who are burdened by the rule would not consider that burden to be a heavy one, then these values suggest that Kaldor-Hicks criteria could be satisfied by a mandatory rule. If it can also be shown that the absence of the mandatory rule creates its own burdens, and especially if these burdens are likely to be heavy to the many who would prefer the result required by the mandatory rule, then it becomes extremely plausible to believe that it is efficient to let that result be a consequence of a mandatory rule.

A. Virtually No Debtor Wants Out of PMSI Superpriority

Professor Buckley asserts that the efficiencies of security interests in after-acquired property are readily apparent. The efficiencies are readily apparent if one is concerned with the relations among a debtor, a secured party and the general creditors. Those efficiencies are not, however, apparent in the relation among debtor, the first-in-time secured party and second-in-time secured party. Analysis suggests that there are very few debtors who would freely consent to eliminating by contract their capacity to create PMSI superpriority.

Behind the UCC's after-acquired property rule lies the assumption that once a secured creditor takes a security interest in property, the

debtor's productive activity will produce revenues which could be used either to repay the creditor or buy more property. The after-acquired property allows the debtor to offer substitutes for the original collateral without having to go through the formality of paying off the original loan and taking out a new loan. Thus, within the UCC, the debtor's use of his own funds is made the equivalent of using a new advance from the first secured lender.[18] The transactional efficiencies arising by means of the automatic perfection of the secured party's interest in the after-acquired collateral are obvious. An enormous amount of paperwork tracing the flow of funds from the debtor to the creditor and back to the debtor, is eliminated.[19]

PMSI superpriority arises only in circumstances in which the source of funds for purchasing the after-acquired property is not the debtor, but a second secured creditor. The question, then, is what are the circumstances that would cause a debtor to choose to go to a second secured lender rather than either returning to the first secured lender or using his own assets to replace the security originally provided to the first lender.

While it is possible that the second creditor could give the debtor a better deal than the first creditor, it is by no means obvious how this circumstance would come about. Both first and second secured creditors sell their funds in extremely competitive capital markets, and presumably have similar investment opportunitites. Thus, it is unlikely that there is something about second secured creditors that give them inherently cheaper money. If there were, one would wonder why they were not first secured creditors.

It is possible that the advantage of the second secured creditor is simply that she is second-in-time. To the extent that the first secured creditor has search or investigative costs on which the second secured creditor can free-ride, that saving could account for the capacity of second secured creditors to do better by debtors who have granted security

---

18. To the extent that the source of funds is an unsecured creditor, so long as ostensible title issues can be assumed to be solved, for purposes of this discussion there is no difference between a debtor using his own funds or borrowing them from a knowledgeable general creditor and paying the appropriate risk premium.

19. Whether this transactional efficiency is also Kaldor-Hicks efficient cannot be said with certainty unless one assumes continuous full employment. The claim of efficiency does not take into account the welfare of the small army of clerks that the adoption of the UCC left unemployed. A further Kaldor-Hicks analysis would have to ask whether, with the increase in volume of financing activity that the UCC created, enough of this group became reemployed. If so, then further analysis would require evaluating alternative life-time income flows of these clerks, with or without the UCC, measured against the gains to financial institutions, as well as the presumed reduction in prices paid by ordinary consumers as a consequence of a lower price of credit.

interests in their after-acquired property. The UCC provides ample opportunitites for second secured creditors to take advantage of first secured creditors in this way. Thus, one would expect that in loan contracts now entered into under the UCC there would appear some type of protection for the first secured creditor against this risk. These defensive measures to costs that mandatory PMSI superpriority would appear to impose on the first secured lender are likely to include prepayment penalties or negative pledge agreements. Professor Scott, whose theory of secured transactions predicted that these types of clauses would appear whenever creditors created blanket liens (including after-acquired property clauses), went out to look for them and could not find them. Their absence, as he freely admitted, put the validity of his theory into question.[20] Their absence also makes this explanation of second-in-time secured creditor's advantage unlikely.

If these search and investigative costs must be incurred by any creditor, and, if it is assumed that the second secured creditor does not take advantage of the first secured creditor's information, then the capacity of the second secured creditor to give the debtor a better deal than the first secured creditor becomes even more mysterious. When the debtor seeks a new loan, the first-in-time creditor's sunk costs ought to enable her to make that loan to the debtor at rates lower than any second-in-time creditor, since she need not incur costs to acquire information she already has.

Yet the reality remains that second-in-time secured creditors can and do give better deals to debtors than do first-in-time secured creditors. The principal source of the second-in-time secured creditor's advantage must lie in the collateral itself. The second-in-time secured creditor, when judging the value on default of the collateral she finances, must place a higher value on that collateral than does the first-in-time secured creditor. Evidence within the UCC permits the conclusion that the drafters believed that financing sellers had special capacities to realize value in the event they repossessed the goods they sold, and they systematically provided statutory advantages for those sellers. There is also substantial evidence that the drafters had good reason for their belief.

B. PMSI Superpriority within the UCC

1. The Statutory Provisions.—The very complexity of the provisions within the UCC creating PMSI superpriority suggests some of the reasons for having PMSI superpriority. In the contest between later PMSI holders

---

20. Scott, *supra* note 3, at 951. Scott also reports that the equivalent of mortgage points, a device familiar to anyone who has mortgaged the family home, is also virtually unknown in personal property security.

and first-in-time secured creditors with after-acquired property rights, both Article 2 and Article 9 have something to say. Within Article 9, section 9-312 is the principal arbiter. If the collateral is inventory, subsection (3) gives superpriority to the PMSI lender, provided she gives proper notice to the first-in-time secured party whose interest in the same collateral has been created by an after acquired property clause.[21] In collateral "other than inventory," subsection (4) gives superpriority to the PMSI, provided the PMSI lender perfects within 10 days of the date when the debtor "receives possession" of the collateral.[22] The rights that Sections 2-702 and 2-705 give to sellers of goods have the effect of creating a PMSI in goods sold. The rights under Section 2-705 are superior to those of a first-in-time secured creditor with rights to after-acquired property, while PMSI rights under Section 2-702 are subordinated to a secured creditor with rights in after-acquired property.[23]

---

21. U.C.C. § 9-312(3) states:

> A perfected purchase money security interest in inventory has priority over a conflicting security interest in the same inventory and also has priority in identifiable cash proceeds received on or before the delivery of the inventory to a buyer if
>
> (a) the purchase money security interest is perfected at the time the debtor receives possession of the inventory; and
>
> (b) the purchase money secured party gives notification in writing to the holder of the conflicting security interest if the holder had filed a financing statement covering the same types of inventory (i) before the date of the filing made by the purchase money secured party, or (ii) before the beginning of the 21 day period where the purchase money security interest is temporarily perfected without filing or possession (subsection (5) of Section 9-304); and
>
> (c) the holder of the conflicting security interest receives the notification within five years before the debtor receives possession of the inventory; and
>
> (d) the notification states that the person giving the notice has or expects to acquire a purchase money security interest in inventory of the debtor, describing such inventory by item or type.

22. U.C.C. § 9-312(4) reads:

> A purchase money security interest in collateral other than inventory has priority over a conflicting security interest in the same collateral or its proceeds if the purchase money security interest is perfected at the time the debtor receives possession of the collateral or within ten days thereafter.

23. Under section 2-702, a seller has a right against an insolvent buyer to reclaim goods "upon demand made within ten days after the receipt." But this right of reclamation is "subject to the rights of a buyer in the ordinary course of business or other good faith purchaser under this Article (Section 2-403)." A holder of a perfected security interest in the same goods would qualify as a "good faith purchaser under this Article." *See* Gilmore, *The Good Faith Purchase Idea and the Uniform Commercial Code: Confessions of a Repentant Draftsman*, 15 Ga. L. Rev. 605, 616 ff. (1981). Under section 2-705, the seller may stop delivery of "carload, truckload, planeload or larger shipments of express or freight when the buyer repudiates or fails to make a payment due before delivery or if for any other reason the seller has a right to withhold or reclaim the goods." This

PMSI lenders do not have to accept the superpriority the statute offers them. In the case of inventory, a PMSI lender can simply choose not to give the required notice, and in any event, for every sort of collateral, Section 9-316 authorizes "any person entitled to priority" to subordinate her priority by agreement.

Despite the expansive wording of Section 9-312(4), PMSI superpriority rights under this subsection are effectively limited to tangible collateral. In his discussion of Section 9-312(4), Grant Gilmore characterizes his discussion of the possibility of PMSI interests arising in general intangibles as being "almost on a hypothetical level."[24] Gilmore's characterization, if anything, overstates the possibility of applying Section 9-312(4) to intangibles, since he did not consider the effect of the requirement that the PMSI lender perfect before or within 10 days of when the debtor "receives possession" of the collateral. The metaphysical difficulties involved in receiving possession of anything intangible are formidable.[25]

When a secured lender relies on priority rights to after-acquired intangibles, her expectations may be defeated by making those rights tangible and then having the tangible version of those rights acquired by a good faith purchaser. If, for example, a lender relies on a pool of continuously renewing accounts receivable, a debtor could defeat the creditor's interest by having his customers embody their obligations in negotiable instruments. The debtor could then sell these instruments to a person who qualifies as a holder in due course, leaving the original secured creditor without assets she legitimately expected to be available to her. Consistent with his views concerning mandatory PMSI superpriority, Professor Buckley would abolish negotiability, at least to the extent it has the consequence of subordinating after-acquired property financers. The real problem is, of course, the risk of debtor dishonesty.[26]

---

section does not offer any general protection to the secured party as a good faith purchaser. *See* A. Farnsworth & J. Hunnold, Commercial Law, Cases and Materials 720-50 (4th ed. 1985).

24. 2 G. Gilmore, Security Interests in Personal Property § 29.5 (1965).

25. Gilmore limits his discussion of this requirement in the statute to a one sentence cross reference to the parallel language in section 9-312(3). *See id.* at 799. With respect to inventory, section 9-312(3)'s reference to "receives possession" creates no mysteries.

26. *Cf.* Buckley, *supra* note 2, at 1467. A debtor who under the UCC would convert accounts receivable into tangible paper, would, under Professor Buckley's alternative regime, be likely to forge records of accounts receivable. Nor is it so certain that because "the market for negotiable instruments . . . has largely disappeared," negotiability has lost its importance for financers. *Id.* at 1469. One use for negotiability in commercial financing is in bankers acceptances. The Federal Reserve reports that for the six months of April to September of 1986, the volume of bankers acceptances was approximately $12 billion per month, of which approximately $2 billion per month consisted of drafts

2. *One Policy Rationale for the UCC Provision.*—There can be no doubt that creditors, when examining a debtor's tangible property, consider the costs of repossessing and selling that property. Creditors will assign different values to the property. If a creditor has property assigned to her, one component of that valuation has to be the creditor's own capacity to resell the goods. If all other things are equal, a rule of law that assigns the property to the highest valuing creditor will do only what a debtor would choose to do himself.[27] One creditor who is highly likely to have advantages over other creditors in disposing of property will be the creditor who first sold it to the debtor on credit. There is evidence that this type of repossessing seller can on resale sometimes realize an amount in excess of the original contract price.[28] That seller is likely to have her greatest advantage if the goods are still in the same form and packaged for transport in the same way as that seller had left them. Not surprisingly, Article 2 creates a PMSI in favor of reclaiming sellers and reserves its strongest security interest for the creditor-seller of goods who stops them while they are in transport.

---

purchased by the holding bank. 73 FED. RES. BULL. A23 (Jan. 1987). While this $2 billion is admittedly a small fraction of the commercial paper and bankers acceptance market, which the same source puts at close to $400 billion per month, it does seem an overstatement to say that a $2 billion market has "largely" disappeared.

It is also worth noting that a substantial number of existing financial practices assume the existence of negotiable instruments. Among the most important of these practices is the modern check clearing system. This is not to say that check clearing could be accomplished without negotiability, but only to say that the current system does depend on negotiability in important ways. *See* Rogers, *The Irrelevance of Negotiable Instruments Concepts in the Law of the Check-Based Payment System*, 65 TEX. L. REV. 929 (1987). Thus, even if Professor Buckley's intuitions were correct and the "case for holder in due course superpriority" were "speculative" enough to justify the conclusion that if we were planning the world anew, we would not create this superpriority, the transition costs that would be incurred to abolish it from our world do act as an argument against this reform. Given our starting point, it is not enough to show that having no holder in due course superpriority would be a better state of affairs than having it. One must show that the improvement would be worth enough to pay the costs of the transition as well.

27. The highest valuing creditor will not (except for certain accounts receivable transactions) be the economists' familiar highest valuing user of a particular good. Since voluntary creditors are by definition individuals who have entered into transactions with the debtor, had they been the highest valuing user of the collateral, they would have bought it from the debtor. The qualification that transaction costs could have prevented such a sale is always present and possibly true. However, since the creditor and debtor were already engaged in a loan transaction in which the debtor had informed the creditor of the existence of the items standing as collateral, it is hard to imagine how transaction costs are likely to inhibit significant percentages of transactions that have reached this point of information exchange.

28. *See* Shuchman, *Profit on Default: An Archival Study of Automobile Repossession and Resale*, 22 STAN. L. REV. 20 *passim* (1969).

This analysis is sufficient to explain why the UCC creates PMSIs arising by reason of law under Article 2. The sellers benefitted by these provsisions are likely to be in a better position than any other creditors to dispose of the goods to best advantage. While creditors as a class would prefer to have the goods, even at the price of an additional claimant in the seller's bankruptcy, the overall wealth of this particular unhappy community is most probably increased by creditors generally foregoing their claim to goods and having the original seller dispose of the goods for the best price possible. The economic justification for that imputed creditor preference is supported by a story made familiar by agency costs analysis.[29]

The implicit efficiency account concerning sellers becomes less compelling the more the seller's goods are commingled with the debtor's general assets. There is still reason to believe that the seller will have advantages over general creditors in disposing of the goods, but that advantage is not so probable if the competitor for the goods is a secured creditor who has chosen to create an interest in the same assets. Here *a priori* economic analysis has little to say, other than that the outcome depends on the relative advantages of the two secured creditors in disposing of the goods. The inconclusiveness of the economic analysis does, however, suggest why Section 2-702 has remained a major subject for debate.[30]

Under Article 9 of the UCC, any secured creditor can achieve superpriority over secured creditors by following the procedures set forth in Sections 9-312(3) or 9-312(4). The creditor who obtains superpriority under either of these subsections is likely to be a creditor who can anticipate that her valuation of the collateral will exceed that which would be given by the secured party of record. The statutory arrangements permit the debtor to take advantage of the higher valuing creditor's opinion. The superpriority provisions permit this transaction without disturbing the pre-existing credit relationship.[31]

---

29. The standard agency cost story applies here. The seller, so long as he captures the entire profit on resale, will have an incentive to make the investment that yields the maximum price. To the extent the seller shares those profits with others, his incentive to make investment necessary to obtain the maximum price will be correspondingly reduced. *See generally,* A. Barnea, R. Haugen, L. Senbet, Agency Problems and Financial Contracting (1985), and especially Chapter III, *The Nature of Agency Problems.*

30. *See* J. White & R. Summers, Uniform Commercial Code, 1027-28 (2d. ed. 1980).

31. Why a second secured creditor would value his collateral more highly than the first secured creditor may be appreciated in the context of specific cases. One example is the case of Brown and Williamson Tobacco Corp. v. First Nat'l Bank of Blue Island, 504 F.2d 998 (7th Cir. 1974). The debtor was a cigarette wholesaler who bought and resold cigarettes made by most of the major tobacco companies. The branded cigarettes

The plausible perceptions about relations of secured creditors to collateral that underlie the provisions in both Article 2 and Article 9 suggest strongly that in a regime of free contract virtually all debtors would insist on PMSI superpriority. To show that PMSI superpriority would be extremely common does not, however, mean that it ought to be a mandatory rule. To make that judgment, Professor Buckley's attack on PMSI superpriority requires an examination of the costs to those contracting parties who do not wish to have PMSI superpriority and compare that cost to the relative drafting burdens under regimes with either mandatory PMSI superpriority or free contract. The question remains whether anything sensible can be said regarding the question of who should bear the drafting burden—the PMSI lender wanting superpriority or the first-in-time secured party wanting superior rights in after-acquired property.

C. *Weighing the Costs of the Drafting Burden*

1. *Silence as Consent to What?*—Mandatory PMSI superpriority, although wanted by most contracting parties, may not be wanted by everyone. Whether the benefits to the large number of people wanting the statutory contract do outweigh the unavoidable costs to those not wanting that contract is very much an empirical question. The data to answer it does not, and in all probability cannot, exist. There are, however, a few considerations which, in the absence of data, suggest that if the data did exist, the data would support arrangements similar to those now provided by the UCC.

The question at issue is what happens to contracting parties if they fail to agree explicitly to create PMSI superiority. The default rule in a regime of free contract would create a result that most contracting parties would not want. The great majority of contracting parties in such a regime would have provided for PMSI superiority by contracts containing standard language. There will, however, always be a few people who, because they do not consult a lawyer or because they consult an incompetent lawyer, end up with contracts that do not contain the standard language. This circumstance creates what Soia Mentschikoff once called the "poor schnook" problem. These people will have contracts that fail to express their intent. Some judges will view the failure to include the standard language as conclusively expressing the parties' intent to contract contrary to the norm. A sympathetic court, however, might well observe both the common occurrence of the language and the absence in the case at hand of the special circumstances that usually

---

obviously had a special value after repossession to the original manufacturer that they did not have to the other manufacturers.

explain for the absence of the language. The court could then to choose to impose as a matter of law the legal consequences of the usual language. This process (often observable when the default rule does not match the ordinary practice) creates a mandatory regime piecemeal. Since it is piecemeal, neither the default rule nor its opposite can be relied upon with anything approaching certainty.[32] Since Professor Buckley's free contract regime's default rule would be contrary to the desires of most contracting parties, it would be especially vulnerable to the "poor schnook" problem.

2. *Search Costs and the Drafting Burden.*—Professor Buckley finds reduction in search costs to be a sufficient justification for mandatory rules within Article 9, especially those which require filing for perfection of security interests.[33] Assuming that Professor Buckley is correct in his conclusion that the mandatory filing provisions of Article 9 meet his test for validity, these same provisions, to some extent, justify mandatory PMSI superpriority. Debtors and secured creditors find PMSI superpriority simplifies features of the mandatory filing system. Professor Buckley's failure to connect the two parts of his argument weakens the case he makes against mandatory PMSI superpriority.

No mandatory search system *necessarily* reduces search costs. The system will do so if it has identified the information creditors will want to know with some accuracy, and does not burden those creditors with unnecessary information. In addition, the system must not impose substantial costs on those individuals who must provide the data. The UCC's notice filing system is built on the assumption that creditors are interested in classes of potential collateral and that the creditor will view notice that at least some of a debtor's property of a certain class is potentially subject to a secured claim as useful information. This method of information categorization makes dealing with a floating lien on all of the debtor's property, whether existing or after-acquired, somewhat difficult.[34] The burden on a creditor desiring to create such a floating lien, however, is not very great.

---

32. *See, e.g.*, A. Trollope, Phineas Redux, CH. LX "Two Days Before the Trial" Trollope introduces the character of Chaffanbrass, the best criminal defense barrister in England. In this passage, he is in conversation with another attorney. Chaffanbrass says, "Caveat emptor is the only motto going, and the worst proverb that ever came from dishonest stoney-hearted Rome. With such a motto as that to guide us no man dare trust his brother. Caveat lex—and let the man who cheats cheat at his peril." "You'd give the law a great deal to do." "Much less than at present. What does your Caveat emptor come to? That every seller tries to pick the eyes out of the head of the purchaser. Sooner or later the law must interfere, and Caveat emptor falls to the ground.

33. *See* Buckley, *supra* note 2, at 1453-56.

34. *See, e.g.*, cases noted in J. White & R. Summers, Uniform Commercial Code

If the assumption concerning what creditors want to learn from a search is correct, then the notice provisions in the UCC succeed in reducing search costs. A quick look at a financing statement informs interested creditors whether further investigation will be needed. If the system is efficient, the answer will often be no. Because the UCC encourages thinking about collateral in broad categories, it also has the effect of encouraging both debtors and secured parties to create claims of security interests that go beyond what truly concerns the secured creditor. A secured creditor interested in only one part of a debtor's inventory will, by reason of U.C.C. Section 9-402, give notice to the world that she has an interest in the debtor's inventory. While it is true that the security agreement can limit the creditor's claim, the legal climate in existence at the time the UCC was written, and which still exists, would encourage a cautious creditor to make expansive claims in the security agreement as well. The UCC does not give creditors and debtors any substantial motive for drafting security agreements narrowly. Since security agreements may be amended, no second-in-time secured party would comfortably rely on a security agreement which makes a narrow claim to part of a debtor's inventory where the financing statement asserts the first-in-time creditor has rights to "inventory." A sensible second-in-time secured creditor without PMSI superpriority would want to communicate with the first-in-time secured creditor to enter into an express agreement limiting the first creditor's claim.[35]

To the extent that efficient creditor search strategies encourage broad descriptions of the claims to collateral, both in the security agreements and in the financing statements, these efficient strategies will result in apparently overreaching claims by the first-in-time secured creditor. Mandatory PMSI superpriority counterbalances this type of overreaching drafting. If there were no mandatory PMSIs, then first-in-time secured lenders and their debtors would have substantial reasons to write much more precise descriptions, both in filings and in security agreements. More precise descriptions in security agreements and financing statements would be especially needed for inventory security arrangements. Because

---

910 n.50 (2d. ed. 1980). The point is not that these cases are correct, but simply that the existence of these cases mean the risk to secured parties with overbroad descriptions in their security agreements is real.

35. Under Professor Buckley's alternative system, any second-in-time secured creditor relying on an arrangement between the debtor and a first-in-time secured creditor which prevents the first-in-time secured creditor from claiming superpriority in after-acquired property would also have similar fears concerning amendment of that contract. Any cautious PMSI lender claiming superpriority under that alternative system would want to contract with the first-in-time secured party. Thus, under that system, a transaction that the UCC accomplishes with one contract would require two contracts.

debtors and creditors often expect that the original seller will replace the original inventory as it turns over, these arrangements are likely to require after-acquired property clauses. When, however, the debtor has multiple sources for his inventory, (*i.e.*, where the debtor is a retailer selling a broad range of products each supplied by a different financing manufacturer) the need for the debtor's business flexibility combined with the continuing creditors' need for after-acquired protection against competing secured parties will necessarily lead debtors and creditors to draft security agreements with limited and precise descriptions of the collateral pledged to each creditor.

These more precise descriptions would, in a perverse way, have the effect of increasing the search costs of any second-in-time secured creditor who wished to enter into what today would be a mandatory PMSI superpriority transaction. Unless the first-in-time secured creditor had, by contract, permitted the creation of PMSI superpriority, the second-in-time creditor would have to determine the precise scope of the first-in-time secured party's claim to the collateral. That investigation alone would have some cost, as well as always creating the residual risk of error by the second-in-time secured creditor. Nor would these more limited descriptions of collateral necessarily be helpful to the first-in-time secured creditor. Since the debtor now would have an incentive to insist on narrowly drafted grants of security interests, the first-in-time secured creditor would have an increased risk that after default, property she thought was subject to the security interest would be determined by a court not to be.[36]

Compared to the UCC, Professor Buckley's alternative regime carries with it a high probability of greater search costs for second-in-time secured lenders and, at least for inventory first-in-time secured lenders, a higher probability of risk of catastrophic losses due to hostile court interpretations of limiting language. To believe that Professor Buckley's alternative regime would be preferable to the UCC's structure, one has to believe that the sorts of costs that this regime would create for both first and second-in-time secured parties would be less than the value to debtors of having the ability to create, with certainty, rights of first-in-time secured parties to after-acquired property. In the absence of data, no belief can be unequivocably declared unreasonable. In light of the risks that one can identify in Professor Buckley's alternative regime, and in light of the difficulty in identifying the circumstances in which

---

36. Considering the frequency with which this rather unpleasant surprise occurs under the UCC, where the debtor does not have the same motive to create narrow grants of security interests, this risk of finding no security for what had been made as a secured loan does appear to be more than academic.

debtors would want to contract with first-in-time secured creditors without at the same time reserving the right to create PMSI superpriority, it does appear reasonable to conclude that the burden of proof still remains on Professor Buckley.

D. The Risk of Debtor Dishonesty

Professor Buckley considers and rejects the argument that, under existing law, debtors who wish to offer first-in-time secured creditors the benefit of after-acquired property clauses could do so by contract. He notes that even though "an after-acquired property loan agreement might declare the grant of a subsequent PMSI interest to be an event of default, once made and registered, the PMSI interest is effective as against the first lender."[37] Debtor dishonesty also creates troubling issues for his proposed alternative to existing law.

In his preferred regime of free contract, the risk of debtor dishonesty falls entirely on the second-in-time PMSI lender. A dishonest debtor could fraudulently create a security agreement in which the first-in-time secured creditor appears to give the debtor rights to create PMSI superpriority. Public recording by the first-in-time secured creditor of notice that the debtor had rights to create PMSI superpriority would reduce opportunities for this sort of debtor misbehavior, but not completely eliminate it. A dishonest debtor could still enter into a recorded security agreement forbidding PMSI superpriority and then forge a document amending the recorded instrument to permit him to enter into PMSI superpriority credit arrangements.

Under existing law, first-in-time secured creditors know they cannot totally rely on having a monopoly on after-acquired property. They can, however, expect to have both rights to after-acquired property and substantial protection against debtor dishonesty. This protection may be found in the structure of the transaction. Since UCC Section 9-312(3) requires a would-be PMSI lender to give notice to the prior secured party in order to obtain PMSI superpriority, the first-in-time secured creditor can use that notification to force a default and immediate repayment, having taken the (reasonable under all circumstances) precaution of not having the debt exceed the amount of the security. With respect to equipment and other tangible non-inventory security, a first-in-time secured creditor fearful of debtor dishonesty can arrange the payment schedule so that the amount outstanding is at all times equal to no more than a realistic depreciation schedule for the item standing as security.[38]

---

37. Buckley, *supra* note 2, at 1465 n.163.

38. Any secured creditor, whether under the existing or any proposed alternative regime, would of course have to monitor to be sure that the original collateral had not been disposed of in untraceable ways.

Under the free contract regime, the second-in-time secured creditor who has been tricked into believing she has superpriority rights has no obvious means of protecting herself in the structure of her transaction. Instead, the second-in-time secured creditor must find her protection by entering into direct communications with the first-in-time secured creditor because no second-in-time creditor can safely rely on whatever documents either the debtor or public records produce. The second-in-time secured creditor will have to ask the first-in-time secured creditor to describe her rights to the debtor's after-acquired property. In every case where the UCC requires one communication, there would be two communications under Professor Buckley's alternative system.

In addition, second-in-time secured creditors will have to enter into two-way communications under circumstances where the UCC requires none. This communication would confirm that the first-in-time secured party has no rights and does not intend to have rights. Even where first-in-time secured parties intend to have rights in after-acquired property, debtors will likely respond by insisting on narrowly drawn security agreements. In these cases, cautious second-in-time secured parties must be sure that the first-in-time secured party's claim is subject to the limits stated in the documents possessed by the debtor. First-in-time secured creditors are not likely to admit to limited claims without involving lawyers in the communication process, which of course increases the costs of communication.

When the concern is debtor dishonesty, the question becomes whether the costs to the second-in-time secured creditors whose debtors want PMSI superpriority under a free contract regime will exceed the costs under existing statutes to those first-in-time secured creditors who want after-acquired property superpriority. Again, in the absence of data no certain conclusions are possible. It is possible, however, to identify the interests which would have to be compared. Under Professor Buckley's alternative system, second-in-time secured lenders would have to invest to acquire information sufficient to avoid the total disaster which would result if a dishonest debtor were wrongfully to convince a creditor that he had the power to create PMSI superpriority. Under the UCC, suspicious first-in-time secured creditors will incur the costs of restructuring the transaction to avoid being vulnerable to debtor dishonesty. Since these steps in restructuring the transaction have the result of keeping the value of the security below that of the debt, it is likely that first-in-time secured parties will take these steps even without the fear of debtor dishonesty. Even when first-in-time creditors would not take these sorts of steps, it is reasonable to believe that self-help will be less costly to secured creditors than would be the cost of acquiring the information that insures against disaster. If this reasonable belief is true, then the risks created by debtor dishonesty are less burdensome under the UCC then they would be under the regime of free contract.

IV. Conclusion

Professor Buckley offers a seductively attractive proposition to criticize the UCC's rules concerning PMSI superpriority. When he says that mandatory rules need special justification, he sounds as if he favors freedom of contract over regulation. What he makes us forget for a while is that any legal order, by providing for an outcome in the absence of agreement to the contrary, places burdens on contracting parties. All Professor Buckley would have us do is substitute one rule for another. The rule he proposes has the advantage that if the contracting parties do not like it, they can freely contract around it. His alternative rule, however, has two principal disadvantages. Nearly everyone would choose to contract around it, and no one who has contracted around it could be certain that if the debtor does breach his contract, the law would be able to provide an adequate remedy to the second-in-time secured party, the party presumably benefited by a contract term creating PMSI superpriority.

The case for the efficiency of Professor Buckley's alternative system is not especially plausible. While he is correct to suggest that the UCC does not take into account the negative side to what its drafters saw as transaction cost-reducing mechanisms, a further examination of its provisions suggests that there is reason to believe that these costs are not great. There is even stronger reason to believe that the very few secured creditors who are disadvantaged by mandatory PMSI superpriority would find these costs less than the costs imposed on all secured creditors contracting in the absence of these statutory rules. So long as Kaldor-Hicks criteria provide the basis for a judgment about efficiency, the numbers of persons benefited by the present arrangements, the degree of their benefit, the near-absence of anyone wanting an alternative regime combined with the costs to persons injured by breach of contracts needed in Professor Buckley's proposed alternative regime, all suggest that PMSI superpriority under the UCC is preferable to this proposed alternative.

Random Drug Testing of Police Officers: A Proposed Procedure Which Satisfies Fourth Amendment Requirements

"[F]or nothing is more certain than that beneficent aims, however great or well directed, can never serve in lieu of constitutional power."[1]

INTRODUCTION		799
I.	OVERVIEW OF THE DRUG PROBLEM	801
	A. Drug Abuse as an Immense National Problem	801
	B. Criticisms of Drug Testing	802
II.	CURRENT DRUG TESTING METHODS	804
	A. The EMIT Test	804
	B. The GC/MS Test	805
III.	FOURTH AMENDMENT ANALYSIS	806
	A. Is Urine Testing a Search?	806
	B. The Reasonable Suspicion Standard	807
	C. Heavily Regulated Industries	809
IV.	ANALYSIS OF CASE LAW	810
	A. Reasonable Suspicion Required	810
	B. Reasonable Suspicion Not Required	813
	C. Can the Conflicts be Reconciled?	815
V.	MODEL REGULATION & TESTING PROCEDURE	817
VI.	CONCLUSION	820

INTRODUCTION

The problem of illegal drug use in the work place has received considerable attention in recent years.[2] The public now considers drug abuse more threatening than even real war.[3] In response to this increased awareness, many employers, both public and private, have begun testing for illegal drug use among their employees.[4] In the public sector, the

1. Carter v. Carter Coal Co., 298 U.S. 238, 291 (1936).

2. *See* Weisman, *48 Hours on Crack Street: I Was A Drug-Hype Junkie*, NEW REPUBLIC, Oct. 6, 1986, at 14, 15 (article discusses the vast number of stories on drug abuse).

3. *See America's Crusade: What is Behind the Latest War on Drugs*, TIME, Sept. 15, 1986, at 60; *Presidential Debate* (network television broadcast, Sept. 25, 1988). Illegal drug use is the most important public problem today. *Id.*

4. Stille, *Drug Testing: The Scene is Set for a Dramatic Legal Collision Between the Rights of Employers and Workers*, NAT'L L.J., Apr. 7, 1986, at 1, col. 1. In 1986 about one-fourth of the leading industrial companies conducted drug testing. *Id.* Almost five million Americans were tested for drug use in 1985. *Id.* Fifty percent of Fortune 500 companies now have testing programs. 134 CONG. REC. S7811 (daily ed. June 14, 1988).

military was the first to test on a significant basis.[5] In 1987, the Reagan Administration accelerated the assault on drug use in the work place when the President signed Executive Order 12,564.[6] This executive order implemented drug testing in the federal executive branch.[7] Based on this order, many police departments began testing their members.

Since drug testing of police officers began, the question as to the proper standards required, in terms of inception of testing and scope of testing, has been examined by many courts.[8] Until recently, the courts have consistently required some basis for suspecting that the police officer was using illegal drugs before upholding the legality of the testing procedure.[9] In *Policemen's Benevolent Association v. Washington Township*,[10] a court for the first time failed to require reasonable suspicion before testing police officers.[11] This case is in direct conflict with other recent cases which require reasonable suspicion.[12]

This note will examine the fourth amendment issues encountered in the mandatory[13] drug testing of police officers. Particular emphasis will be paid to the recent case of *Policemen's Benevolent Association*.[14] Part I gives an overview of the drug problem and discusses the goals of any

---

5. Stille, *supra* note 4, at 1, col. 1. The military now conducts about half of all drug tests performed in the nation. *Id.* at 22, col. 2.

6. Exec. Order No. 12,564, 3 C.F.R. 224 (1987).

7. *Id.* Testing is limited to federal employees holding sensitive positions. *Id.* at 226. The executive order defines sensitive position based on the agency's mission, the employee's duties, and the potential danger to the public health, safety, or national security. *Id.*

8. *See infra* notes 89-170 and accompanying text for a discussion of the principal cases.

9. Lovvorn v. Chattanooga, 647 F. Supp. 875, 881 (E.D. Tenn. 1986), 846 F.2d 1539 (6th Cir. 1988). *See infra* notes 90-136 and accompanying text.

10. 850 F.2d 133 (3d Cir. 1988). This case is titled *Policemen's Benevolent Ass'n v. Washington Township* in the District Court decision. 672 F. Supp. 779 (D.N.J. 1987). It is titled *Policeman's Benevolent Ass'n v. Washington Township* in the Court of Appeals decision. 850 F.2d 133 (3d Cir. 1988). However, the Court of Appeals refers to *Policemen's* Benevolent Association in the text. *Id.* at 134. Therefore, this author will use *Policemen's* Benevolent Association throughout this Note when referring to either the opinion of the District Court or Court of Appeals.

11. *Id.* at 141.

12. *See infra* notes 90-136 and accompanying text.

13. In this note the terms random and mandatory drug testing are used interchangeably. Random or mandatory testing refers to any testing where the degree of suspicion is general to the group rather than specific to the individual. Random testing involves selecting persons to be tested by random methods, while mandatory testing involves testing the entire police force. The terms are used interchangeably because the issues involved are identical. *See* Ayers, *Constitutional Issues Implicated by Public Employee Drug Testing*, 14 WM. MITCHELL L. REV. 337, 340-41 (1988).

14. 850 F.2d 133 (3d Cir. 1988).

drug testing program. Part II explores the various testing procedures. Fourth amendment issues are discussed in Part III. Part IV provides an analysis of the recent cases. A sample drug testing procedure which would meet all the fourth amendment constraints is provided in Part V. Finally, Part VI concludes with the assertion that mandatory drug testing of police officers is constitutionally permissible without individualized reasonable suspicion.

I. Overview of the Drug Problem

A. Drug Abuse as an Immense National Problem

Although illegal drug use has been a part of American life for a long time, the problem has increased dramatically in the last twenty years.[15] Reportedly, a quarter of the American population use, either regularly or occasionally, some kind of illegal drug.[16] Drug use in the work place can lead to an increase in property damage, tardiness, absenteeism, employee theft, health care costs, workers' compensation costs, and the number and severity of accidents while decreasing the productivity and workmanship of the employees.[17] American business loses an estimated sixty billion a year because of employee drug use.[18]

Drug use in the public sector is even more problematic than drug use in the private sector. The government has an obligation to the

---

15. Adams, Blanken, Ferguson & Reznikov, Overview of Selected Drug Trends 1 (1985)(available from the National Institute on Drug Abuse). An estimated fifty-six million Americans have used marijuana, twenty million are current users. *Id.* at 11. Twenty-two million Americans have tried cocaine, four million of whom are current users. *Id.* at 7.

16. Stille, *supra* note 4, at 1, col. 1.

17. *See* Geidt, *Drug and Alcohol Abuse in the Work Place: Balancing Employer and Employee Rights*, 11 Emp. Rel. L.J. 181, 181 (1985); Stille, *supra* note 4, at 1, col. 2; 134 Cong. Rec. S7811 (daily ed. June 14, 1988).

18. 134 Cong. Rec. S7811 (daily ed. June 14, 1988). Employees with drugs in their systems are:

two and one-half times more likely to have absences of eight days or more;

three and six-tenths times more likely to injure themselves or another person in a work place accident;

five times more likely to be involved in an accident off the job, which, in turn, affects attendance or performance on the job;

five times more likely to file a workers' compensation claim;

one-third less productive; and,

incur 300 percent higher medical costs and benefits.

In 1988, Congress projected that the federal government would spend more than $2.5 billion in 1988 on interdiction, investigation, prosecution, correction, enforcement, and assistance to state and local governments. *Id.* Congress further projected that local police departments would spend another $861 million. *Id.*

community. Drug use among its employees hinders the satisfaction of this obligation.[19] This duty to the public is particularly critical when the government employee is charged with protecting the public's safety. Police are entrusted with the unusual authority to use force. The potential for disastrous results is obvious if policemen are under the influence of drugs.

The police officer's duties render him especially susceptible to illegal drug use.[20] Many police officers are in constant contact with drug users and suppliers. Most police officers know how to obtain illegal drugs. Another important contributing factor to a police officer's susceptibility to use illegal drugs is the high stress involved in the job. It has even been stated that it would be an anomaly if drug abuse was not a problem among police officers.[21] One author outlines seven reasons to test police officers.[22] The reasons are public safety, public trust, potential for corruption, presentation of credible testimony, morale in the force, loss of productivity, and civil liability.[23]

B. Criticisms of Drug Testing

Even though there are many strong reasons to test for illegal drug use among police officers, many problems are associated with testing. Critics of drug testing claim that drug testing does not significantly increase the safety of the work place and that testing programs have revealed a rather low level of illegal drug use.[24] Proponents point to several incidences where drug use has significantly declined after testing had begun.[25] Clearly drug testing can reduce the negative effects of illegal

---

19. *See supra* note 17 and accompanying text.

20. Dunham, Lewis & Alpert, *Law Enforcement: Testing the Police for Drugs*, 24 Crim. L. Bull. 155 (1988).

21. *Id.* (citing G. Alpert & R. Dunham, Policing Urban America (1988)).

22. Higginbotham, *Urinalysis Drug Testing Programs for Law Enforcement (Part I)*, 55 FBI L. Enf. Bull. 25 (Oct. 1986).

23. *Id. See* Dunham, Lewis & Alpert, *supra* note 20, at 156-58 for a discussion of these seven reasons to test police officers.

24. Of 5,300 people tested by the Customs Service, only six had positive results, about 0.1 percent. Neal, *Mandatory Drug Testing*, 74 A.B.A. J. 58, 63 (Oct. 1988). Out of 1508 railroad workers tested, ten tested positive for alcohol and sixty-six tested positive for drugs, approximately 5.8 percent. *Id.*

25. Within two years of beginning drug testing, the Southern Pacific Railroad experienced a decrease in accidents of seventy-two percent. After instituting a testing program, Armco National Supply Company's plant in Houston experienced a two-thirds reduction in its accident rate, a fifteen percent increase in productivity, an increase in product quality, and a decrease in turnover. The United States Navy's positive rate for sailors was reduced from forty-seven percent to four percent in just three years. The Navy reported a significant increase in battle readiness. 134 Cong. Rec. S7812 (daily ed. June 14, 1988).

drug use if conducted properly. Critics also challenge the reliability of the drug tests and predict that false positive results will ruin the careers of tens of thousands of innocent people.[26] These criticisms are not without merit and must be addressed before random drug testing is allowable. This note proposes a procedure that will provide nearly one hundred percent reliability.[27]

Other potential problems, besides reliability of the actual chemical test, must be addressed in a comprehensive urinalysis procedure. Drug testing is considered to be embarrassing and degrading.[28] Although courts have differed on the level of embarrassment created by drug testing,[29] any embarrassment must be minimized. Another problem is that tests for illegal drugs do not determine the level of impairment at the time of testing.[30] Evidence of the drug use may appear in a urine sample for weeks after taking the drug.[31] The root question is whether an officer can be disciplined for off duty illegal drug use. This question is less difficult to answer with regard to police officers than for other public employees for two reasons. First, impairment can last long after the drug is ingested, and long after the user notices the impairment.[32] Second, the actions of off duty police officers are already regulated by their departments. Police department procedures require that off duty police officers conduct their private lives so as to not bring the department into disrepute.[33] They must carry their revolver and badge at all times.[34]

---

26. Stille, *supra* note 4, at 1, col. 2. *See infra* notes 36-51 and accompanying text.

27. *See infra* notes 171-178 and accompanying text.

28. *See, e.g.*, National Treasury Employees Union v. Von Raab, 816 F.2d 170, 175 (5th Cir. 1987), *aff'd in part*, 109 S. Ct. 1384 (1989); Capua v. City of Plainfield, 643 F.Supp. 1507, 1514 (D.N.J. 1986).

29. "A urine test done under close surveillance of a government representative, regardless of how professionally or courteously conducted, is likely to be a very embarrassing and humiliating experience." *Capua*, 643 F. Supp. at 1514. Some courts compare a urinalysis to a body cavity search. Tucker v. Dickey, 613 F. Supp. 1124, 1129-30 (W.D. Wis. 1985). Another court stated that a urinalysis "is even less intrusive than a fingerprint which requires that one's fingers be smeared with black grease." Mack v. United States, Fed. Bureau of Investigation, 653 F. Supp. 70, 75 (S.D.N.Y. 1986).

30. The level of impairment has not been correlated with specific concentrations of drug metabolites. *See* National Institute on Drug Abuse, Q & A, Detection of Drug Use by Urinalysis 13 (1986).

31. A urine test for marijuana can detect casual use for up to two weeks and even longer for a chronic user. *See* Centers for Disease Control, *infra* note 43, at 469-70.

32. A study completed on pilots using a flight simulator showed impairment even twenty-four hours after smoking marijuana. One pilot, who reported no awareness of any intoxication, completely missed the runway. Yesavage, Leirer, Denari & Hollister, *Carry-Over Effects of Marijuana Intoxication on Aircraft Pilot Performance: A Preliminary Report*, 142 Am. J. Psychiatry 1325, 1328 (1985).

33. Policemen's Benevolent Ass'n v. Washington Township, 850 F.2d 133, 138 (3d Cir. 1988).

34. *Id.* at 139-40.

Therefore, level of impairment at the time of testing is not as important as would be true for other public employees because the actions of police officers can be regulated even if they are off duty. Many opponents of drug testing point out that over-the-counter medications can affect the results.[35] This must be considered in developing a sound testing procedure. Due to these problems, courts have been reluctant to allow police departments to freely test their officers.

II. Current Drug Testing Methods

One of the basic assumptions of any drug testing program is that the individuals who test positive have actually used illegal drugs. Considering the seriousness it is imperative that the results be accurate.

A. *The EMIT Test*

The most common urinalysis test is the enzyme multiplied immunoassay technique (EMIT).[36] The EMIT test can be used to test for marijuana, cocaine, barbiturates, amphetamines, phencyclidine (PCP), and opiates (including heroin).[37]

The EMIT test is the most popular because it is inexpensive,[38] requires little formal training to administer, has a short analysis time, can detect a wide range of drugs, and is promoted by its manufacturer as being ninety-five percent accurate.[39] The EMIT test is also popular because it may be performed at the workplace.[40]

The EMIT test has been criticized because it has a tendency to yield false positive results when legitimate drugs are mixed with certain foods.[41]

---

35. *See infra* note 41 and accompanying text.

36. L. Dogoloff & R. Angarola, Urine Testing in the Workplace 11 (1985); M. Houts, R. Baselt & R. Craven, Courtroom Toxicology §§ 30.01-30.06 (1986).

37. L. Dogoloff & R. Angarola, *supra* note 36, at 21.

38. The cost of an EMIT is approximately $5. The cost of a gas chromatography/mass spectrometry (GC/MS) test is between $25 and $35. Note, *Random Drug Testing of Government Employees: A Constitutional Procedure*, 54 U. Chi. L. Rev. 1335, 1338 n.11 (1987). Confirmation tests can cost up to eighty dollars apiece. Stille, *supra* note 4, at 23, col. 4.

39. *Survey of the Law on Employee Drug Testing*, 42 U. Miami L. Rev. 553, 565 (1988).

40. Morgan, *Problems of Mass Urine Screening for Misused Drugs*, 16 J. Psychoactive Drugs 305, 306 (1984).

41. *See Grapevine*, Time, Oct. 3, 1988, at 25. A study of how over-the-counter medications can affect the results of a urinalysis found that "Advil can produce a positive reading for marijuana; swallowing Alka-Seltzer can lead to a positive verdict of amphetamines; ingesting Vicks cough medicine or Robitussin-DM can yield a positive indication of morphine." *Id. See also* Budiansky, *Busting the Drug Testers*, U.S. News & World Report, Oct. 20, 1986, at 70 (reporting that a poppyseed bagel triggered a positive result

Also, the EMIT test has been criticized because passively inhaled marijuana smoke could register as a positive result.[42] Some experts claim that the EMIT gives inaccurate results sixty percent of the time.[43] Because of the limitations of the test, a manufacturer of the EMIT test has recommended that a positive test result be confirmed using another testing method.[44]

B. The GC/MS Test

The most conclusive confirmation test is the gas chromatography/mass spectrometry test (GC/MS).[45] The GC/MS test is not used as an initial screening device due to its cost[46] and must be performed in a laboratory by trained technicians.[47] The GC/MS test requires more equipment and takes longer than the EMIT test. But the GC/MS test has the advantage of producing a graphic record that an individual expert can review.[48] Since a GC/MS test requires a high level of training to perform,[49] a poorly trained technician could misinterpret the results. But if a GC/MS test is performed carefully by an adequately trained technician, it is almost always accurate.[50]

---

for cocaine); Zeese, *infra* note 42, at 26 (Syva Company, the manufacturer of EMIT, has reported that aspirin, amphetamine, amitriptyline, benzocyclecgonine, diazepam, meperidine, methaqualone, morphine, phencyclidine, propoxyphene, and secobarbital may create false positive results).

42. *See* Zeese, *Marijuana Urinalysis Tests*, 1 Drug L. Rep. 25, 28 (1983) (reporting a study conducted by the *American Journal of Psychiatry* in 1977 that found positive test results after passive inhalation of marijuana smoke).

43. Hansen, Caudill & Boone, *Crisis in Drug Testing*, 253 J. A.M.A. 2382, 2386 (1985). *But see*, Centers for Disease Control, *Urine Testing for Detection of Marijuana Use: An Advisory*, 32 Morbidity & Mortality Weekly Rep. 469 (1987) (urinalysis only four percent inaccurate).

44. Miller, *Mandatory Urinalysis Testing and the Privacy Rights of Subject Employees: Toward a General Rule of Legality Under the Fourth Amendment*, 48 U. Pitt. L. Rev. 201, 205 (1986). Stille, *supra* note 4, at 23, col. 4. Several courts have found that double EMIT testing is adequate to satisfy the due process clause of the Constitution. *See*, Wykoff v. Resig, 613 F. Supp. 1504 (N.D. Ind. 1985); Peranzo v. Coughlin, 608 F. Supp. 1504 (S.D.N.Y. 1985); Spence v. Farrier, 807 F.2d 753 (8th Cir. 1986). Some courts even allow reliance on a single EMIT test. *See* Jensen v. Lick, 589 F. Supp. 35 (D.N.D. 1984)(prison officials could impose sanctions on prisoners based on an unconfirmed EMIT test); Smith v. State, 250 Ga. 438, 298 S.E.2d 482 (1983)(the EMIT test is sufficiently reliable to stand as the only evidence in a parole revocation hearing).

45. Schwartz & Hawks, *Laboratory Detection of Marijuana Use*, 254 J. A.M.A. 788, 790 (1985).

46. *See supra* note 38.

47. L. Dogoloff & R. Angarola, *supra* note 36, at 22.

48. McBay, *Problems in Testing for Abused Drugs*, 255 J. A.M.A. 39, 40 (1986).

49. L. Dogoloff & R. Angarola, *supra* note 36, at 22.

50. National Treasury Employees Union v. Von Raab, 816 F.2d 170, 181 (5th Cir. 1987), *aff'd in part*, 109 S. Ct. 1384 (1989).

Since the tests themselves are nearly one hundred percent accurate, the real potential for error is with the administration of the tests.[51] Sloppy laboratory practices, such as failure to clean the equipment properly, will cause erroneous results. Therefore, a strict chain of custody must be developed.

III. Fourth Amendment Analysis

The fourth amendment protects people from unreasonable searches and seizures undertaken by either federal or state governments.[52] The purpose of the fourth amendment is to "safeguard the privacy and security of individuals against arbitrary invasions by government officials."[53] To analyze drug testing under the fourth amendment, two issues must be addressed. First, is urinalysis a search under the fourth amendment? If so, is that search reasonable and thus exempt from the warrant requirement?[54]

A. Is Urine Testing a Search?

An overwhelming number of federal district courts and federal courts of appeal have held that mandatory taking of urine constitutes a search

---

51. Since laboratories are unregulated, the level of quality varies enormously. Studies have set the error rate at between three and twenty percent. Stille, *supra* note 4, at 24, col. 3.

52. The fourth amendment to the Constitution provides:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

U.S. Const. amend. IV. *See* Smith v. White, 666 F. Supp. 1085, 1089 (E.D. Tenn. 1987).

53. Camara v. Municipal Court, 387 U.S. 523, 528 (1967).

54. On November 2, 1988, the United States Supreme Court heard arguments on two cases concerning drug testing by the public sector. Legal experts noted that the Court stepped in rather quickly to reconcile the lower courts. The Court considered *National Treasury Employees Union v. Von Raab*, 816 F.2d 170 (5th Cir. 1987), *aff'd in part*, 109 S. Ct. 1384 (1989), and *Railway Labor Executives' Ass'n v. Burnley*, 839 F.2d 575 (9th Cir. 1988), *rev'd sub nom., Skinner v. Railway Labor Executives' Ass'n*, 109 S. Ct. 1402 (1989).

National Treasury involved the testing of Customs Service employees when they applied for specific sensitive positions. *Burnley* challenged the constitutionality of administering drug tests to all crew members of a train involved in an accident that results in property damage or personal injury. Neither case involved random drug testing.

In both cases, the Supreme Court held that a urinalysis is a search under the fourth amendment. *National Treasury*, 109 S. Ct. at 1390; *Railway Labor Executives*, 109 S. Ct. at 1413. Also, in both cases the Supreme Court held that the search was reasonable and thus exempt from the warrant requirement. *National Treasury*, 109 S. Ct. at 1397; *Railway Labor Executives*, 109 S. Ct. at 1422. Since neither case involved random drug testing, the constitutionality of this procedure has not been established.

and seizure under the fourth amendment.[55] Courts consider urinalysis a search because of the information about a person's personal life which may be determined.[56] It can be determined whether a person is diabetic, whether a woman is pregnant, and whether a person is taking any prescription drugs.[57] While the Supreme Court has not decided whether the mandatory taking of a public employee's urine constitutes a search and seizure, the Court has held that most bodily intrusions, such as involuntary blood test, are a search and seizure under the fourth amendment.[58] The Court determined that the drawing of blood was a minor intrusion.[59] If the Supreme Court ever considers this issue, the Court will likely hold that mandatory production of urine is a search and seizure under the fourth amendment.[60]

B. The Reasonable Suspicion Standard

Generally, probable cause is required before a search or seizure is considered reasonable.[61] However, courts have held that searches based on a lesser showing can still be reasonable.[62] For a search to be reasonable

---

55. Neal, *supra* note 24, at 60. *See, e.g.*, National Fed'n of Fed. Employees v. Weinberger, 818 F.2d 935, 942 (D.C. Cir. 1987); National Treasury Employees Union v. Von Raab, 816 F.2d 170, 176 (5th Cir. 1987), *aff'd in part*, 109 S. Ct. 1384 (1989); McDonell v. Hunter, 809 F.2d 1302, 1307 (8th Cir. 1987); Capua v. City of Plainfield, 643 F. Supp. 1507, 1513 (D.N.J. 1986); Jones v. McKenzie, 628 F. Supp. 1500, 1508-09 (D.D.C. 1986), *rev'd*, 833 F.2d 335 (D.C. Cir. 1987); Allen v. City of Marietta, 601 F. Supp. 482, 488-89 (N.D. Ga. 1985); Storms v. Coughlin, 600 F. Supp. 1214, 1217 (S.D.N.Y. 1984); Patchogue-Medford Congress of Teachers v. Board of Educ., 119 A.D.2d 35, 505 N.Y.S.2d 888, 889 (1986); City of Palm Bay v. Bauman, 475 So. 2d 1322, 1325-27 (Fla. Dist. Ct. App. 1985).

56. McDonell v. Hunter, 612 F. Supp. 1122, 1127 (S.D. Iowa 1985).

57. Stille, *supra* note 4, at 22, col. 1. A urinalysis also can determine whether one is being treated for a heart condition, manic-depression, epilepsy, or schizophrenia. *Id.*

58. Schmerber v. California, 384 U.S. 757, 767 (1966). Recently, the Supreme Court has held that a mandatory urinalysis is a search under the fourth amendment. *See supra* note 54.

59. *Id.* at 771-72. *See also* Winston v. Lee, 470 U.S. 753, 759 (1985) (a compelled surgical procedure is a major intrusion "of such magnitude that the intrusion may be 'unreasonable,' even if likely to produce evidence of a crime."); Cupp v. Murphy, 412 U.S. 291, 296 (1973) (the taking of scrapings from a murder suspect's fingernails was a minor intrusion).

60. "If at one time it might have been possible to argue that urinalysis does not constitute a search or seizure, such an argument is now entirely untenable." American Fed'n of Gov't Employees v. Weinberger, 651 F. Supp. 726, 732 (S.D. Ga. 1986).

61. O'Connor v. Ortega, 107 S. Ct. 1492, 1501 (1987). *See* Coolidge v. New Hampshire, 403 U.S. 443, 449-50 (1971); United States v. Harris, 403 U.S. 573, 577 (1971); Aguilar v. Texas, 378 U.S. 108, 115 (1964).

62. *See* cases cited *infra* note 72.

with less than probable cause, there must be reasonable ground to believe that the search will result in evidence of drug use and the method must not be excessively intrusive.[63] "[W]hat is reasonable depends on the context within which a search takes place."[64]

The Supreme Court employed a balancing test in *O'Connor v. Ortega*.[65] The Court stated that "[i]n the case of searches conducted by a public employer, we must balance the invasion of the employees' legitimate expectations of privacy against the government's need for supervision, control and efficient operation of the workplace."[66] The Court said a public employer need not have probable cause to conduct such a search, but merely "reasonable grounds for suspecting that the search will turn up evidence that the employee is guilty."[67]

It is difficult to develop a set of rules to determine what is reasonable under the fourth amendment. Justice Scalia, concurring in *O'Connor*, stated that a review of fourth amendment constraints on employer searches provides "a standard so devoid of content that it produces rather than eliminates uncertainty."[68] In each case, a balancing of the necessity of a particular search must be weighed against the invasion of an individual's personal rights.[69] Courts analyze administrative searches different from criminal investigation searches.[70] Administrative searches generally require only the lesser standard of reasonable suspicion. Courts have consistently held that urine testing of public employees falls within the administrative search category.[71] Most of these courts addressing the issue have applied a reasonable suspicion standard to searches that fall under the administrative search exception.[72] Though reasonable suspicion

---

63. Taylor v. O'Grady, 669 F. Supp. 1422, 1436 (N.D. Ill. 1987).

64. New Jersey v. T.L.O., 469 U.S. 325, 337 (1985).

65. 107 S. Ct. 1492 (1987).

66. *Id.* at 1499.

67. *Id.* at 1503. The Supreme Court did not decide whether individual suspicion is an essential element of the standard of reasonableness. *Id.* This failure to require individualized suspicion leaves open the question of the constitutionality of random drug testing under the balancing test. *See infra* text accompanying notes 164-168.

68. O'Connor v. Ortega, 107 S. Ct. 1492, 1505 (1987)(Scalia, J., concurring).

69. Bell v. Wolfish, 441 U.S. 520, 559 (1979).

70. *O'Connor*, 107 S. Ct. at 1501-02.

71. *See* Shoemaker v. Handel, 795 F.2d 1136, 1142 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986); Allen v. City of Marietta, 601 F. Supp. 482, 489 (N.D. Ga. 1985); cases cited *infra* note 72.

72. *See, e.g.*, National Treasury Employees Union v. Von Raab, 109 S. Ct. 1384 (1989); Skinner v. Railway Labor Executives' Ass'n, 109 S. Ct. 1402 (1989); Lovvorn v. City of Chattanooga, 846 F.2d 1539 (6th Cir. 1988); McDonell v. Hunter, 809 F.2d 1302 (8th Cir. 1987); American Fed'n of Gov't Employees v. Weinberger, 651 F. Supp. 726 (S.D. Ga. 1986); Capua v. City of Plainfield, 643 F. Supp. 1507 (D.N.J. 1986); Jones v. McKenzie, 628 F. Supp. 1500 (D.D.C. 1986), *rev'd*, 833 F.2d 335 (D.C. Cir. 1987);

is a lesser standard than probable cause, individualized suspicion is still required.[73]

Courts are split on the level of intrusiveness inherent in a urinalysis. Some courts compare the procedure to body cavity searches[74] while other courts equate the intrusiveness of a urinalysis to that of a blood test.[75] Still other courts hold that a urinalysis is less intrusive than a blood test.[76] The Supreme Court, in *Schmerber v. California*,[77] used a medical procedure's risk, trauma, and pain to judge the intrusiveness of a blood test.[78]

C. *Heavily Regulated Industries*

Generally a warrant is necessary for a search to be considered reasonable under the fourth amendment.[79] Courts have granted an exception to the general warrant requirement when the search involves a heavily regulated industry pursuant to an administrative inspection scheme.[80] These courts have found a diminished expectation of privacy due to the nature of the employee's job.[81] An employee of a heavily regulated industry is used to having his behavior controlled by regulations.[82] Therefore, an additional requirement that the employee submit to a urine test would be less intrusive than for an employee of an unregulated industry. A strong state interest in drug testing is still required.[83] In *Shoemaker v. Handel*[84] the United States Court of Appeals

---

City of Palm Bay v. Bauman, 475 So. 2d 1322 (Fla. Dist. Ct. App. 1985). *See also* Ayers, *Constitutional Issues Implicated by Public Employee Drug Testing*, 14 Wm. Mitchell L. Rev. 337, 345 n.34 (1988).

73. Note, *Drug Testing of Government Employees and the Fourth Amendment: The Need for a Reasonable Suspicion Standard*, 64 Notre Dame L. Rev. 1063, 1073 (1987).

74. *See, e.g.*, Tucker v. Dickey, 613 F. Supp. 1124 (D. Wis. 1985).

75. *See, e.g.*, Shoemaker v. Handel, 795 F.2d 1136 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986); Capua v. City of Plainfield, 643 F. Supp. 1507 (D.N.J. 1986); Storms v. Coughlin, 600 F. Supp. 1214 (S.D.N.Y. 1984).

76. *See, e.g.*, McDonell v. Hunter, 809 F.2d 1302 (8th Cir. 1987); Allen v. City of Marietta, 601 F. Supp. 482 (N.D. Ga. 1985).

77. 384 U.S. 757 (1966).

78. *Id.* at 771. In *Skinner v. Railway Labor Executives' Ass'n*, 109 S. Ct. 1402 (1989), the Supreme Court failed to define the intrusiveness of a mandatory urinalysis.

79. O'Connor v. Ortega, 107 S. Ct. 1492, 1499-1501 (1987).

80. *See, e.g.*, Shoemaker v. Handel, 795 F.2d 1136, 1142 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986); Donovan v. Dewey, 452 U.S. 594 (1981)(warrantless inspection of coal mines); United States v. Biswell, 406 U.S. 311 (1972) (gun selling); Colonnade Catering Corp. V. United States, 397 U.S. 72 (1970) (liquor industry).

81. *See Shoemaker*, 795 F.2d at 1142.

82. *Id.*

83. *Id.*

84. 795 F.2d 1136 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986).

for the Third Circuit allowed the drug testing of jockeys without individual reasonable suspicion under the heavily regulated industry exception.[85] The court held that the heavy regulation of horse racing in New Jersey diminished the jockey's reasonable expectations of privacy and justified a random testing program.[86] *Shoemaker* has been somewhat controversial. Some courts have called it an enigma[87] while other courts have cited it favorably.[88] The next section will analyze recent cases which involve the constitutionality of random drug testing of police officers.

IV. Analysis of Case Law

A. Reasonable Suspicion Required

Until recently, in all cases where courts have ruled on the validity of urine tests for police officers, the courts have used a reasonable suspicion standard as an exception to the warrant requirement, and thus, required individual reasonable suspicion of illegal drug use prior to testing.[89] A thorough analysis of these cases requires an understanding of the particular facts. These facts will be used to analyze other decisions.

In *Capua v. City of Plainfield*[90] the Federal District Court, District of New Jersey, held mass drug testing of police officers to be an unreasonable search and seizure in violation of the fourth amendment.[91] The court balanced the government's interest in a drug-free police department against the intrusiveness of the urine testing procedure.[92] In this case the police officers had no prior notice of the city's intent to conduct mass drug testing.[93] There was no written directive or regulation establishing the basis for testing and defining the testing procedures.[94] A female police employee was tested under the surveillance of a testing

---

85. *Id.* at 1142.

86. *Id.*

87. *See, e.g.*, American Fed'n of Gov't Employees v. Weinberger, 651 F. Supp. 726, 734 (S.D. Ga. 1986).

88. *See, e.g.*, McDonell v. Hunter, 809 F.2d 1302, 1308 (8th Cir. 1987).

89. *See supra* notes 61-78 and accompanying text.

90. 643 F. Supp. 1507 (D.N.J. 1986).

91. *Id.* at 1520. The action in *Capua* was instituted by the City of Plainfield Fire Department and a separate action was instituted by a single female police officer of the City of Plainfield Police Department. *Id.* at 1512. Both the Plainfield Police Department and the Plainfield Fire Department were subjected to similar urine testing at approximately the same time. *Id.* Therefore, the court in *Capua* considered the separate actions jointly. *Id.* Accordingly, the opinion is applicable to both police and fire departments.

92. *Id.* at 1513-20.

93. *Id.* at 1511-12.

94. *Id.* at 1511.

agent of the same sex.[95] After testing positive, a police officer was informed that she could resign without charges being brought or she would be immediately suspended.[96] The court first determined the taking of a urine specimen constitutes a search and seizure within the meaning of the fourth amendment because "each individual has a reasonable expectation of privacy in the personal 'information' bodily fluids contain."[97] The court then examined the unreasonableness of the search. While granting that the city has a legitimate concern with the ability of the police officers to do their duties, the court determined that the tests were intrusive.[98] The court specifically condemned surveillance during collection,[99] the failure to establish procedures to protect the collateral private personal medical information divulged,[100] the lack of confidentiality of the test results,[101] and the lack of prior warning before testing.[102] Lastly, the court held that the fourth amendment allows testing only on a basis of individual reasonable suspicion, not mere suspicion, predicated upon specific facts.[103]

In *Penny v. Kennedy*[104] department-wide urine tests were administered to all members of the police force after four to six days prior notice.[105] Police department officials did not witness the donations of the first half of the tests, but did witness the last half of the tests.[106] The urine tests were initiated based on rumors of switched urine samples in a previous test, an unconfirmed tip from the Federal Bureau of Investigation that an officer had been in contact with a drug dealer, and a statement by an officer that several officers used marijuana.[107] Citing *Lovvorn*, the court required individual reasonable suspicion before the testing of police officers for illegal drugs.[108] Based on the facts presented, the court found "no reasonable suspicion which would justify the administration of these tests at this time."[109] Therefore, the tests constituted

---

95. *Id.* at 1512.
96. *Id.*
97. *Id.* at 1513.
98. *Id.* at 1516.
99. *Id.* at 1514.
100. *Id.* at 1515.
101. *Id.*
102. *Id.*
103. *Id.* at 1517-18.
104. 648 F. Supp. 815 (E.D. Tenn. 1986), *aff'd*, 846 F.2d 1563 (1988). *See* Lovvorn v. Chattanooga, 647 F. Supp. 875 (E.D. Tenn. 1986), *aff'd*, 846 F.2d 1539 (1988). *Lovvorn* opinion attached as an appendix to the *Penny* opinion.
105. 648 F. Supp. at 816.
106. *Id.*
107. *Id.*
108. *Id.* at 817.
109. *Id.*

an illegal search and seizure in violation of the fourth amendment.[110] The court emphasized that drug testing of police officers was possible, but the tests "must be given on reasonable suspicion, their scope must be related to their objective, and they must not be excessively intrusive."[111]

In *Bostic v. McClendon*[112] the police chief was informed by a fellow officer that some members of the police force were smoking marijuana.[113] Without prior notice a urinalysis test was conducted on all members of the police force.[114] The officers were told to submit to the test or be terminated.[115] The urinalysis was to detect marijuana use and not other drugs.[116] The court first held that the taking of a urine specimen is a seizure within the meaning of the fourth amendment.[117] After noting that a search conducted after voluntary consent does not violate the fourth amendment,[118] the court found that where employment would have been terminated if the officers did not participate in the testing, the consent was not voluntary, but was the result of coercion.[119] Therefore, the search and seizure must pass the balancing test to be constitutional. In conducting the balancing test, the court stated that the police officers "do have a significant expectation of privacy which has been infringed upon by the government conduct at issue"[120] while recognizing the strong interest of the police department in protecting the public.[121] The court stressed the authority and discretion provided each officer and the inherently dangerous environment in which they work.[122] The court held that "the [f]ourth [a]mendment allows the [police department] to demand a urine sample from an employee for chemical analysis only on the basis of a reasonable suspicion, based on specific objective facts and reasonable inferences from those facts in the light of experience, that a urinalysis will produce evidence of illegal drug use by that particular employee."[123]

In *City of Palm Bay v. Bauman*[124] the police department notified all officers that consumption of non-prescribed drugs at any time is

---

110. *Id.*
111. *Id.*
112. 650 F. Supp. 245 (N.D. Ga. 1986).
113. *Id.* at 247.
114. *Id.* at 248.
115. *Id.*
116. *Id.*
117. *Id.* at 249.
118. *Id.* (citing Schneckloth v. Bustamonte, 412 U.S. 218, 248-49 (1973)).
119. *Id.*
120. *Id.* at 250.
121. *Id.*
122. *Id.*
123. *Id.*
124. 475 So. 2d 1322 (Fla. Dist. Ct. App. 1985).

strictly prohibited and that urine samples may be required on a random basis.[125] Officers who tested positive would be subject to discipline or discharge.[126] Officers were required to sign this notice.[127] All members of the police department were instructed to submit to urine tests to determine if they had used marijuana.[128] Failure to submit to testing would result in discipline up to and including termination.[129] In determining the reasonableness of the search, the court concluded that police officers, because of the nature of their jobs, should have a diminished expectation of privacy.[130] Police officers must reasonably expect their employer to have a legitimate concern that their ability to discharge their responsibilities is not compromised by drug use because they use weapons, drive vehicles and make instant life and death judgments.[131] Credibility and public confidence is necessary to complete their duty to the public. In balancing the intrusiveness of the search against the public interest, the court considered that the Chief of Police had received no information that any member of the department had used marijuana.[132] The department policy did not articulate any standards for the implementation of the testing program.[133] After considering these facts, the court held that the search and seizure was constitutionally unreasonable and that the police officer's signing of the notice did not constitute consent because the signatures were coerced.[134] The court stated that a city has the right to prohibit police officers from using controlled substances at any time, whether such use is on or off the job.[135] In order to test police officers, the city must have an individual reasonable suspicion based on "specific objective facts and rational inferences that they are entitled to draw from these facts in the light of their experience."[136]

B. Reasonable Suspicion Not Required

The federal district courts and state courts have consistently struck down ill-conceived, reactionary drug testing plans instituted due to the

---

125. *Id.* at 1323.
126. *Id.*
127. *Id.* at 1324.
128. *Id.* at 1323.
129. *Id.*
130. *Id.* at 1324.
131. *Id.*
132. *Id.* at 1325.
133. *Id.*
134. *Id.*
135. *Id.* at 1326.
136. *Id.*

public awareness of illegal drug use.[137] Consequently, many public employers have revised their drug testing procedures to include constitutional safeguards.[138] The courts now must determine whether the revised procedures are adequate.

In *Policemen's Benevolent Association v. Washington Township*,[139] the United States Court of Appeals for the Third Circuit became the first court not to require reasonable suspicion under the fourth amendment before random drug testing police officers.[140]

In *Policemen's Benevolent Association* the mayor notified the police officers that the township would begin a mandatory drug testing program.[141] This announcement contained no details of the proposed plan.[142] Based on this notice, the police officers filed a lawsuit challenging the constitutionality of mandatory drug testing.[143] Thereafter, the township formulated a drug testing procedure.[144] This procedure called for both testing based on reasonable suspicion and random testing.[145] The plan also required an annual medical examination, including a urinalysis drug test.[146] The plan provided that urination would take place in private, unless there was reasonable suspicion that the testee would tamper with the sample.[147] Further, the officers would be notified sixty days in advance of testing and strict testing procedures would be followed to protect the confidentiality of the results.[148] This plan also provided for a drug abuse education program.[149] The police officers then challenged the constitutionality of that plan.[150] The court held that, based on these facts, random drug testing fell within the heavily regulated industry exception to the fourth amendment warrant requirement because the pervasive regulation of the police department reduced the expectation of privacy.[151]

The court based its decision on *Shoemaker v. Handel*.[152] In *Shoemaker*, the Third Circuit Court of Appeals held that the administrative

---

137. *See supra* notes 89-136 and accompanying text.

138. *See, e.g.*, Policemen's Benevolent Ass'n v. Washington Township, 850 F.2d 133, 134 (3d Cir. 1988).

139. 850 F.2d 133 (3d Cir. 1988).

140. *Id.*

141. *Id.* at 134.

142. *Id.*

143. *Id.*

144. *Id.*

145. *Id.* at 135.

146. *Id.*

147. *Policemen's Benevolent Ass'n*, 672 F. Supp. 779, 782 (D.N.J. 1987), *rev'd*, 850 F.2d 133 (1988).

148. *Id.*

149. *Id.*

150. *Policemen's Benevolent Ass'n*, 850 F.2d at 134.

151. *Id.* at 141.

152. 795 F.2d 1136 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986). Other courts

search exception to the fourth amendment applies to the warrantless testing of persons engaged in a highly regulated activity.[153] *Shoemaker* concerned the testing of jockeys.[154] The heavily regulated industry exception has two requirements. First, there must be a strong public interest in conducting a warrantless search, and second, the many regulations of the industry must reduce the justifiable privacy expectation of the individual.[155] The court in *Policemen's Benevolent Association* noted that police officers are probably the most highly regulated of all state employees.[156]

C. *Can the Conflicts be Reconciled?*

Upon initial examination of the drug testing cases involving police officers, the cases seem to fall into either of two distinct categories. One category holds that individual reasonable suspicion is required before a mandatory urinalysis can be allowed.[157] The other category relies upon the heavily regulated industry exception to not require prior individual reasonable suspicion.[158] But, upon careful examination of the particular facts in each case, it becomes clear that the courts are not far apart.

The courts which have required individual reasonable suspicion use a balancing test.[159] This test balances the public interest in conducting the search against the intrusiveness of the search.[160] If the intrusiveness is greater that the public good, the search is unreasonable and therefore, unconstitutional under the fourth amendment. The heavily regulated industry exception uses the same two criteria as the balancing test. The heavily regulated industry exception requires that there be a strong public interest in conducting the search and that the individual's expectation of privacy be reduced.[161] The "heavily regulated industry exception" courts look at the reduction of the police officer's expectation of privacy

---

have followed *Shoemaker*: *see, e.g.*, National Treasury Employees Union v. Von Raab, 816 F.2d 170 (5th Cir. 1987)(custom service officers), *cert. granted*, 108 S. Ct. 1072 (1988); McDonell v. Hunter, 809 F.2d 1302 (8th Cir. 1987)(correctional officers); Rushton v. Nebraska Pub. Power Dist., 653 F. Supp. 1510 (D. Neb. 1987)(nuclear power plant employees).

153. *Shoemaker*, 795 F.2d at 1142.

154. *Id.* at 1137.

155. *Id.* at 1142.

156. Policemen's Benevolent Ass'n v. Washington Township, 850 F.2d 133, 141 (3d Cir. 1988).

157. *See supra* notes 89-136 and accompanying text.

158. *See supra* notes 137-154 and accompanying text.

159. *See supra* notes 89-136 and accompanying text.

160. O'Connor v. Ortega, 107 S. Ct. 1492 (1987).

161. Shoemaker v. Handel, 795 F.2d 1136, 1142 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986).

from the objective person and the "reasonable suspicion" courts examine the quantum level of intrusiveness, but both courts are examining privacy.

There are several factors to be considered when examining the privacy concerns. Courts have mentioned that prior notice of the testing policy and procedures are important.[162] The thoroughness of the procedures are considered. The procedures should allow for the protection of the identity of the officer and any collateral private medical information. Generally, courts have found surveillance during collection to be quite intrusive.[163]

Although the court in *Policemen's Benevolent Association* used the heavily regulated industry exception,[164] it could have reached the same result under the balancing test. In *McDonell v. Hunter*,[165] the United States Court of Appeals for the Eighth Circuit used a balancing test to uphold random drug testing of prison employees who have inmate contact.[166] The court balanced the intrusiveness of drug testing against the state's interest in prison security.[167] The court also noted the prison employees' diminished expectations of privacy because of the sensitive nature of their jobs.[168] Since the court noted that the state's interest in prison security was at least as strong as the integrity of the horse racing industry, the *McDonell* court seemed to indicate that it could have reached the same result by applying the heavily regulated industry exception.[169]

Of the five principal cases discussed previously,[170] the testing method in *Policemen's Benevolent Association* was the least intrusive. The township had developed a formal testing procedure which protected the privacy of the officers as much as possible.[171] The facts of the cases requiring individual reasonable suspicion reveal testing procedures which did not respect the officers' privacy. The importance of the particular facts to the results of the case is revealed in *Capua*. The *Capua* court stated that perhaps the most critical distinction between the search in their case and the search in *Shoemaker* "is the very careful procedural protections built into the *Shoemaker* testing system and the complete absence

---

162. *See supra* notes 93, 94, 100, 102, 114, and 133 and accompanying text.

163. *See supra* notes 99 and 111 and accompanying text.

164. Policemen's Benevolent Ass'n v. Washington Township, 850 F.2d 133, 135-36 (3d Cir. 1988).

165. 809 F.2d 1302 (8th Cir. 1987), *modifying* 612 F. Supp. 1122 (S.D. Iowa 1985).

166. *Id.* at 1308.

167. *Id.*

168. *Id.* at 1306.

169. *Id.* at 1308.

170. *See supra* notes 90-136 and accompanying text.

171. Policemen's Benevolent Ass'n v. Washington Township, 672 F. Supp. 779, 781-82 (D.N.J. 1987).

of procedural safeguards in [the *Capua*] urinalysis program. The jockeys in *Shoemaker* were assured that the results of their tests would be published only to a very few Commissioners."[172] From this statement it can be inferred that the *Capua* court may have followed *Shoemaker* if greater procedural safeguards were present.

Since all courts agree that there is a large public interest in testing police officers for illegal drug use, the constitutionality of the search is dependant upon the procedural safeguards present to protest the privacy of the officer. The next section contains a model procedure which attempts to provide protections to the police officers required under the fourth amendment, while still accomplishing the goals of the testing program.

V. Model Regulation & Testing Procedure

This model regulation and testing procedure attempts to provide adequate safeguards against excessive intrusiveness while still being a reasonably effective testing procedure. The concerns expressed by various courts are addressed in the following model testing procedure.[173]

Model Regulation

1. No officer, employee, or official of the police department shall use any controlled dangerous substance or any prescription drug, unless such substance was obtained directly, or through a valid prescription, from a licensed physician. The term "controlled dangerous substance" shall not include distilled spirits, wine, malt beverages, tobacco, or tobacco products.
2. Upon thirty days notice of the ratification of this procedure, every officer, employee, or official of the police department may be subject to a urine test at the direction of the Chief of Police or his designee based on the selection procedure discussed in section 3. Failure to submit to the test will result in suspension from the police department.

---

172. Capua v. City of Plainfield, 643 F. Supp. 1507, 1520 (D.N.J. 1986).

173. These provisions are modeled after the procedures used by the state of New Jersey for testing jockeys. *See* Shoemaker v. Handel, 795 F.2d 1136, 1138 (3d Cir. 1986), *cert. denied*, 479 U.S. 986 (1986). *See also* Customs Directive 51250-02, Drug Screening Program (August 4, 1986) (testing procedure used by the United States Customs Service. Sample forms are included in this directive). *Cf.* Note, *Drug Testing in the Workplace: A Legislative Proposal To Protect Privacy*, 13 J. Legis. 269, 288 (1986). This note makes a legislative proposal for the drug testing of private sector employees. The procedure allows only very limited testing based on prior individual reasonable suspicion. Random testing is prohibited. It should be noted that the public interest in testing private sector employees is minimal when compared with the public interest in a drug free police force.

3. The Chief of Police or his designee may direct that any officer, employee, or official of the police department be tested if individual reasonable suspicion exists.[174] The Chief of Police or his designee may also direct that up to ten percent of the officers, employees, and officials be tested in any sixty day period. The selection of the ten percent to be tested must be on a random basis.[175] No person shall be tested more than twice in any twelve month period, unless individual reasonable suspicion exists.
4. Any officer, employee, or official who is requested to submit to a urine test shall provide the urine sample, without undue delay, to a professional tester working for the Chief of Police or his designee. The tester may observe the providing of the urine sample only if there is individual reasonable suspicion to believe the person being tested will attempt to switch the sample.[176] A visual observation must be conducted by a person of the same sex as the individual being tested. Individual reasonable suspicion that a person is using illegal drugs is individual reasonable suspicion that a person will attempt to switch the samples.
5. The urine sample shall be immediately sealed and tagged on the form provided. Evidence of such sealing shall be indicated by the signature of the tested individual and the tester. The portion of the form which is provided to the laboratory shall not identify the tested individual by name.
6. Control samples must be intermingled with those of the employees being tested. Results of the tests performed on the control samples must be available to any individual who tested positive.
7. The initial analysis of the urine sample shall be made using an enzyme multiplied immunoassay test (EMIT). If the urine sample tests positive, the urine sample must be reexamined using a gas chromatography/mass spectrometry (GC/MS) test. A positive result based on a GC/MS test constitutes a positive result under this procedure.[177]
8. A positive result shall be reported in writing by the testing laboratory directly to the Chief of Police or his designee. The Chief of Police or his assignee will then determine the individual's name by matching the

---

174. *See supra* notes 61-78 and accompanying text.

175. An example of an acceptable procedure would be to assign a number to each person eligible to be tested and then select the numbers to be tested by a random method. A neutral observer should be present.

176. Protection against substitution of urine samples can be provided through means other than direct visual surveillance. These methods include: (1) checking the temperature of the sample, (2) having the individual provide a second sample (it is unlikely that an individual will carry two substitute samples), (3) testing without warning or opportunity to retrieve a stored substitute sample, and (4) having the tester listen to, but not observe, the providing of the urine sample.

177. *See supra* notes 45-51 and accompanying text.

code to the confidential reference list. After determining the individuals who tested positive, the Chief of Police or his designee shall proceed as follows:

a. He shall, as quickly as possible, in writing notify the individual involved.

b. For an individual's first violation, the Chief of Police or his designee shall issue a written reprimand and the individual shall be required to complete an approved rehabilitation program. The individual must be notified in writing as to the consequences of any further violations.

c. For an individual's second violation, the Chief of Police or his designee shall suspend the individual from the police department. Upon successful completion of an approved rehabilitation program, the individual shall be reinstated. It shall be the individual's responsibility to provide written notice that he has successfully completed the rehabilitation program. Failure to complete the rehabilitation program shall result in the termination of the individual. The individual must be notified in writing as to the consequences of a further violation.

d. For an individual's third violation, the Chief of Police or his designee shall permanently terminate the individual from the police force.

9. Before an individual is terminated, the individual must be notified of the charges against him. The Chief of Police or his designee shall give the individual: (1) opportunity to show any error that may exist; (2) the names and nature of the testimony of witnesses against the individual; (3) a meaningful opportunity to be heard in his own defense; and (4) a hearing before an impartial administrator prior to termination.[178]

10. Any information received in the process of testing, including but not limited to medical information, the results of the tests, or any reports or notices filed as a result of these regulations, shall be treated as confidential.[179] Access to the information shall be limited to the Chief of Police or his designee, counsel to the police department, the individual involved, and counsel for the individual involved.[180]

11. Any reports or memos prepared under these regulations shall be stored in a secure area for a period of one year. After one year, the reports or memos shall be destroyed. However, the reports or memos on individuals who have violated this regulation may be maintained for

---

178. *See* Note, *Yellow Rows of Test Tubes: Due Process Constraints on Discharges of Public Employees Based on Drug Urinalysis Testing*, 135 U. Pa. L. Rev. 1623, 1650-55 (1987). This article discusses the procedural due process aspects of drug testing. The note focuses on the employer's actions in response to a positive result. Recommendations made in this article have been incorporated into the model regulation.

179. *Id.* at 1650.

180. *See supra* note 176.

the purpose of recording the number of violations and the results of rehabilitation, and for use should future violations occur.
12. The test results may not be used for the purpose of criminal prosecution.

VI. Conclusion

Mandatory drug testing of police officers can aid the government in implementing the laudable goals of public safety and confidence in the police force and can enable the government to better fulfill its responsibilities. Since 1987, many police departments have begun mandatory testing programs. Questions developed as to the proper standards for inception of testing and scope of testing.

The early testing programs provided inadequate protection of the police officer's privacy. Due to the lack of responsible testing by the police departments, the programs were determined to be unconstitutional. Individual reasonable suspicion was required before an officer could be tested.

In *Policemen's Benevolent Association v. Washington Township*,[181] a court for the first time upheld random drug testing of police officers. The court based its decision, in part on the procedural protections provided in the testing program.

In order to meet their duty to the public, police departments must begin to institute testing procedures. A procedure which will meet fourth amendment requirements is provided in Part V of this note. A properly administered random testing program can pass fourth amendment scrutiny.

David B. Boodt

---

181. 850 F.2d 133 (3d Cir. 1988).

Delimiting the Manufacturer's Liability: An Examination of Loss of Consortium Recovery in Strict Products Liability Actions Under Section 402A of the Restatement (Second) of Torts

I. Introduction

In 1963, the judiciary first recognized the manufacturer's strict tort liability in *Greenman v. Yuba Power Products, Inc.*[1] The American Law Institute adopted section 402A of the Restatement (Second) of Torts—Special Liability of Seller of Product for Physical Harm to User or Consumer in 1965.[2] Based largely on policy considerations,[3] strict products liability evolved with great emphasis on consumer protection.[4] This led to the continued expansion of the scope of the doctrine.[5] Yet the courts rarely conducted a detailed analysis of whether the policy goals of section 402A were actually promoted by the ever-expanding scope of the manufacturer's liability.[6]

---

1. 59 Cal. 2d 57, 377 P.2d 897, 27 Cal. Rptr. 697 (1963).

2. Restatement (Second) of Torts § 402A (1965).

3. *See, e.g.*, Greenman, 59 Cal. 2d at 63-64, 377 P.2d at 901, 27 Cal. Rptr. at 701. *See generally* Prosser, *The Fall of the Citadel (Strict Liability to the Consumer)*, 50 Minn. L. Rev. 791 (1966) [hereinafter *Fall of the Citadel*]; Prosser, *The Assault Upon the Citadel (Strict Liability to the Consumer)*, 69 Yale L.J. 1099 (1960) [hereafter *Assault Upon the Citadel*].

4. *See* Vandall, *Our Product Liability System: An Excellent Solution to a Complex Problem*, 64 Den. U.L. Rev. 703, 715 (1988). *See also* Manuel & Richards, *Economic Loss in Strict Liability—Beyond the Realm of Section 402A*, 16 Mem. St. U.L. Rev. 315 (1986).

5. *See* Shepard v. Alexian Bros. Hosp., 33 Cal. App. 3d 606, 612, 109 Cal. Rptr. 132, 135 (1973) (California cases expanded the scope of the *Greenman* doctrine by imposing strict liability on retail dealers (Vandermark v. Ford Motor Co., 61 Cal. 2d 256, 391 P.2d 168, 37 Cal. Rptr. 896 (1964)); wholesale and retail distributers (Barth v. B.F. Goodrich Tire Co., 265 Cal. App. 2d 228, 71 Cal. Rptr. 306 (1968)); home builders (Kriegler v. Eichler Homes, Inc., 269 Cal. App. 2d 224, 74 Cal. Rptr. 749 (1969)); bailors and lessors of personal property (McClaflin v. Bayshore Equipment Rental Co., 274 Cal. App. 2d 446, 79 Cal. Rptr. 337 (1969)); and licensors of chattels (Garcia v. Halsett, 3 Cal. App. 3d 319, 82 Cal. Rptr. 420 (1970)). The standard of strict liability has been held to apply to a defect in design as well as a defect in manufacture (Pike v. Frank G. Hough Co., 2 Cal. 3d 465, 467 P.2d 229, 85 Cal. Rptr. 629 (1970)) and extends not only to actual consumers or users, but to any human being to whom an injury from the defect is reasonably foreseeable (Elmore v. American Motors Corp., 70 Cal.2d 578, 451 P.2d 84, 75 Cal. Rptr. 652 (1969)).

6. Shepard v. Superior Court, 76 Cal. App. 3d 16, 27, 142 Cal. Rptr. 612, 619 (1977) (Kane, J., dissenting).

The recognition of loss of consortium as a redressable cause of action within the context of strict products liability provides an excellent example of the expansion of section 402A with a lack of appropriate judicial discourse. Courts have consistently permitted the loss of consortium claim in strict products liability actions. Upon reviewing the decisions in many jurisdictions, a federal district court noted that it had "not found a single case where a consortium claim was dismissed in a products liability action. To the contrary, the decisions do not even question the validity of a consortium claim."[7] Yet liability under section 402A of the Restatement is strict liability in tort[8]—culpability of the defendant is not a necessary factor. Accordingly, the scope of the manufacturer's responsibility should be carefully delineated.

Whether the loss of consortium cause of action is appropriately within the scope of section 402A liability depends primarily upon policy considerations[9] which have significantly shifted in the two decades since the establishment of strict tort liability. Therefore, this note details the relevant historical aspects of the loss of consortium and strict tort liability doctrines, as well as society's current concern with delimiting liability. Analysis of the apposite case law and Restatement provisions follow. Although valid arguments support both views of the issue, countervailing policies are now sufficient to preclude continued manufacturer responsibility for this legal wrong. Thus, thorough judicial analysis of the loss of consortium claim under a strict products liability theory is both timely and appropriate.

II. Relevand Historical Considerations

A. The Loss of Consortium Cause of Action

A cause of action for loss of consortium has existed for hundreds of years.[10] Significantly, the nature of the claim continually adapted to meet the changing needs of society. Originally, the term consortium denoted the husband's legal right to the wife's performance of the duties and obligations assumed by her upon marriage.[11] The husband's cause

---

7. Timms v. Verson Allsteel Press Co., 520 F. Supp. 1147, 1151 (N.D. Ga. 1981).

8. Restatement, *supra* note 2, § 402A comments a and m.

9. *See* Borer v. American Airlines, Inc., 19 Cal. 3d 441, 563 P.2d 858, 138 Cal. Rptr. 302, 306 (1977) (in delimiting the extent of a tortfeasor's responsibility for damages under the general law of torts, the courts must locate a line between liability and nonliability at some point, a decision which is essentially political).

10. This legal right was first recognized in Guy v. Livesey, 79 Eng. Rep. 428 (1681).

11. *See generally*, Note, *Loss of Consortium: Paradise Lost, Paradise Regained,* 15 Cumb. L. Rev. 179 (1984) (authored by Nancy C. Osborne).

of action originated at a time in history when the wife was wholly subserviant and the law viewed his loss primarily in terms of the deprivation of her services—her performance in caring for the home, rearing the children, and serving her husband.[12] Although the cause of action was first permitted only against one who intentionally infringed upon those rights,[13] the courts ultimately extended the loss of consortium claim to cases in which the injury to the wife was negligently inflicted.[14]

A wife, however, did not originally have a similar cause of action. Prior to the passage of the married woman's acts, a married woman did not have a separate legal identity and could not sue in her own name.[15] The emancipation statutes, therefore, played an important role in the evolution of the loss of consortium action. Once endowed with rights equal to her husband's, the courts quickly recognized the wife's cause of action where the defendant's wrong was an intentional one.[16] However, not until 1950, in *Hitaffer v. Argonne Co.*,[17] did a court allow a wife's claim for loss of consortium due to a negligent injury to her husband.[18] Further, no significant trend followed this decision until ten years later when one court aptly noted that "the obstacles to the wife's action were 'judge invented'" and were therefore subject to judicial destruction.[19] By the 1970's, an overwhelming majority of states recognized the wife's cause of action for loss of consortium.[20]

Concurrent with the grant of the wife's right to claim loss of consortium, a gradual shift in the focus of loss of consortium ensued. Rather than the tangible and pecuniary entitlement to services, the intangible relational aspects of the loss were emphasized.[21] Thus, the modern definition of consortium tends to be vague and indefinite and

---

12. *Id.* at 184.

13. *See* Guy v. Livesey, 79 Eng. Rep. 428 (1681). *See generally* Note, *supra* note 11.

14. *See, e.g.*, Skoglund v. Minneapolis St. Ry., 45 Minn. 330, 47 N.W. 1071 (1891), *overruled*, Boland v. Morrill, 275 Minn. 496, 148 N.W.2d 143 (1967); Busch v. Busch Constr., Inc., 262 N.W.2d 377 (Minn. 1977).

15. *See generally* Chused, *Late Nineteenth Century Married Women's Property Law: Reception of the Early Married Women's Property Acts by Courts and Legislatures*, 29 Am. J. Legal Hist. 3 (1985); Chused, *Married Women's Property Law: 1800 - 1850*, 71 Geo. L.J. 1359 (1983).

16. *See, e.g.*, Hinnant v. Tidewater Power Co., 189 N.C. 120, 126 S.E. 307 (1925), *overruled*, Nicholson v. Hugh Chathan Mem. Hosp., 300 N.C. 295, 266 S.E.2d 818 (1980).

17. 183 F.2d 811 (D.C. Cir. 1950), *overruled on other grounds*, Smither & Co. v. Coles, 242 F.2d 220 (D.C. Cir. 1957).

18. *See Hitaffer*, 183 F.2d at 813.

19. Dini v. Naiditch, 20 Ill. 2d 406, ____, 170 N.E.2d 881, 892 (1960).

20. *See* Note, *supra* note 11, at 190-91 n.85.

21. *Id.*

the "various attempts at defining it only aggrevate its nebulosity."[22] Further, today the term is applied to many different relationships and is defined in the context of the particular relationship for which protection is sought.[23] The definitions typically include the term "society," which one court noted as referring to a "broad range of mutual benefits each family member receives from [the] other's continued existence, including love, affection, care, attention, companionship, comfort and protection."[24]

Accordingly, the loss of consortium claim evolved into the area of filial relationships in recent years. In 1975, in *Shockley v. Prier*,[25] the Wisconsin Supreme Court permitted parents to recover for the loss of their child's aid, comfort, society and companionship resulting from the defendants' negligence.[26] In 1980, in *Ferriter v. Daniel O'Connell's Sons Inc.*,[27] a child's cause of action for the negligent deprivation of parental consortium was granted by the Massachusetts Supreme Court.[28] Moreover, the court in *Ferriter* stated that "as claims for injuries to other relationships come before us, we shall judge them according to their nature and force."[29]

In conclusion, the doctrine of loss of consortium dramatically evolved through the years to conform to contemporary society. Although the courts expanding the cause of action routinely recognized the fear that

---

22. Hodges v. Johnson, 417 S.W.2d 685, 691 (Mo. Ct. App. 1967).

23. *See, e.g.*, Berger v. Weber, 411 Mich. 1, 303 N.W.2d 424 (1981) (child sought recovery for loss of parental society and companionship); Kailimai v. Firestone Tire & Rubber Co., 87 Mich. App. 144, 273 N.W.2d 906 (1978) (spouse sought recovery for loss of society, companionship, services, and all other incidents of the marriage relationship); Wangen v. Ford Motor Co., 97 Wis. 2d 260, 294 N.W.2d 437 (1980) (parent sought recovery for loss of child's services, society, companionship, and pecuniary support); Elden v. Sheldon, 46 Cal. 3d 267, 758 P.2d 582, 250 Cal. Rptr. 254 (1988) (cohabitant sought recovery for loss of conjugal society, comfort, affection, companionship and sexual relations because there was a stable and significant relationship parallel to a marital relationship).

24. Consolidated Machines., Inc. v. Protein Prods. Corp., 428 F. Supp. 209, 228 (M.D. Fla. 1976) (an action for loss of society under maritime laws).

25. 66 Wis. 2d 394, 225 N.W.2d 495 (1975).

26. *Id.* at ____, 225 N.W.2d at 501. *Accord* Reben v. Ely, 146 Ariz. 309, 705 P.2d 1360 (1985); Norvell v. Cuyahoga County Hosp., 11 Ohio App. 3d 70, 463 N.E.2d 111 (1983). *But see* Baxter v. Superior Court, 19 Cal. 3d 461, 563 P.2d 871, 138 Cal. Rptr. 315 (1977); Wilson v. Galt, 100 N.M. 227, 668 P.2d 1104 (1983).

27. 381 Mass. 507, 413 N.E.2d 690 (1980).

28. *Id.* at 413 N.E.2d at 696. *Accord* Weitl v. Moes, 311 N.W.2d 259 (Iowa 1981), *overruled*, Audubon-Exira Ready Mix, Inc. v. Illinios Cent. Gulf R.R., 335 N.W.2d 148 (Iowa 1983); Berger v. Weber, 411 Mich. 1, 303 N.W.2d 424 (1981); Hay v. Medical Center Hosp., 145 Vt. 533, 496 A.2d 939 (1985); Ueland v. Reynolds Metals Co. (State Report Title: Veland v. Pengo Hydra-Pull Corp.), 103 Wash. 2d 131, 691 P.2d 190 (1984); Theama v. City of Kenosha, 117 Wis. 2d 508, 344 N.W.2d 513 (1984).

29. *Ferriter*, 381 Mass at ____, 413 N.E.2d at 696.

liability might ultimately reach too far, the arguments advanced for restraining the claim were generally rejected due to policy considerations, or as one court noted, "due to an absence of sufficient countervailing policy."[30] As a result, it is reasonable to assert that the cause of action will continue to expand.

The cause of action for loss of consortium was initially available when the injury was intentionally inflicted, and only gradually extended to cases of negligence. Research yielded no case adopting a new cause of action for loss of consortium in which the legal injury resulted from conduct governed by the law of strict liability. Because the courts have analyzed and endorsed the validity of loss of consortium recovery in the negligence context, the issue becomes whether the policies and rationales underlying strict tort liability are distinct enough to justify a different result.

B. The Manufacturer's Strict Liability in Tort

Unlike the doctrine of loss of consortium, strict products liability is of relatively recent origin.[31] *MacPherson v. Buick Motor Co.*,[32] decided in 1916, is often noted as the father of modern products liability actions. *MacPherson's* abrogation of the privity requirement in negligence actions[33] was a major development in the evolution of the doctine. Yet the difficult task of proving the negligence of an often remote manufacturer still confronted the plaintiff.[34] Thus, the courts often resorted to the doctrine of res ipsa loquitur which permitted at least an inference of negligence from the presence of a defective product on the market.[35]

A leading re ipsa case, *Escola v. Coca-Cola Bottling Co.*,[36] provided the opportunity for Justice Traynor to enunciate the concept of strict

---

30. Ekalo v. Constructive Serv. Corp., 46 N.J. 82, 215 A.2d 1, 8 (1965). *See also* Swartz v. United States Steel Corp., 293 Ala. 439, ____, 304 So. 2d 881, 883-87 (1974) (the court rejected each of the six arguments contending that allowance of the wife's claim would: (1) ignore the state constitution; (2) repudiate the doctrine of *stare decisis*; (3) violate the doctrine of separation of powers by judicial invasion of the legislative function; (4) create the possibility of double recovery for a single injury; (5) allow an extension of causes of actions to all persons who suffer a loss when a loved one has been injured; and (6) cause difficulties in assessing damages).

31. For a history of products liability, *see generally* R. Epstein, Modern Products Liability Law (1980); Prosser, *Assault Upon the Citadel*, *supra* note 3.

32. 217 N.Y. 382, 111 N.E. 1050 (1916).

33. *Id.* at ____, 111 N.E. at 1053.

34. *See* Prosser, *Assault Upon the Citadel*, *supra* note 3, at 1114.

35. *See, e.g.*, Gordon v. Aztec Brewing Co., 33 Cal. 2d 514, 532, 203 P.2d 522, 524-35 (1949). *See generally* Jaffe, *Res Ipsa Loquitur Vindicated*, 1 Buffalo L. Rev. 1, 13 (1951).

36. 24 Cal. 2d 453, 150 P.2d 436 (1944).

liability in tort in a concurring opinion issued in 1944. Traynor stated that "it should now be recognized that a manufacturer incurs an absolute liability when an article that he has placed on the market, knowing that it is to be used without inspection, proves to have a defect that causes injury to human beings."[37] In justification, Traynor noted the realities of modern manufacturing and marketing as well as the manufacturer's potential to increase product safety and the availability of insurance.[38] He presented a strong argument for the idea that the manufacturers, as a group and an industry, should absorb the inevitable losses which result from the use of their products, because they are in a better position to do so.[39] Thus, it was public policy which prompted the manufacturer's liability.

A number of legal scholars agreed with Traynor's reasoning and urged the judicial creation of strict tort liability.[40] However, no American court adopted such a rule until 1963 in *Greenman v. Yuba Power Products, Inc.*,[41] when Justice Traynor wrote for a unanimous court. The court in *Greenman* held that "[a] manufacturer is strictly liable in tort when an article he places on the market, knowing that it is to be used without inspection for defects, proves to have a defect that causes injury to a human being."[42] Further, the court noted the policy that the costs of injuries resulting from defective products should be borne by the manufacturers who market them rather than by the consumers who are powerless to protect themselves.[43] Other jurisdictions rapidly adopted the "loss shifting" theory enunciated in *Escola* and adopted in *Greenman*. The courts manifested a unified purpose—to protect the consumer[44]—and thus the strict tort liability concept developed with the goal of making it easier for consumers to obtain compensation.[45]

The rapid acceptance of *Greenman* was aided by the American Law Institute's adoption of section 402A of the Restatement (Second) of

---

37. *Escola*, 24 Cal. 2d at 461, 150 P.2d at 440 (Traynor, J., concurring).

38. *Id.* at ____, 150 P.2d 440-41.

39. *See* Prosser, *Assault Upon the Citadel*, *supra* note 3, at 1120.

40. *See, e.g.*, Green, *Should the Manufacturer of General Products Be Liable Without Negligence?*, 24 Tenn. L. Rev. 928 (1957); James, *General Products—Should Manufacturers Be Liable Without Negligence?*, 24 Tenn. L. Rev. 923 (1957); Noel, *Manufacturers of Products—The Drift Toward Strict Liability*, 24 Tenn. L. Rev. 963 (1957); Prosser, *Assault Upon the Citadel*, *supra* note 3; Wilson, *Products Liability Part 1: The Protector of the Injured Person*, 43 Calif. L. Rev. 614 (1955).

41. 59 Cal. 2d 57, 377 P.2d 897, 27 Cal. Rptr. 697 (1963).

42. *Greenman*, 59 Cal.2d at 60, 377 P.2d at 900, 27 Cal. Rptr. at 700.

43. *Id.*

44. *See* Vandall, *supra* note 4, at 715.

45. *Id.* at 710.

Torts[46] two years later. The new section provided for the "special liability of sellers of products for physical harm to users or consumers."[47] However, the drafting of section 402A began well before the judicial recognition of strict tort liability in *Greenman*. In 1961, the Reporter for the Restatement, Dean William Prosser, introduced the original version of section 402A to the American Law Institute (ALI).[48] This draft indicated that a seller would be strictly liable for "bodily harm"[49] resulting from the sale of impure food for human consumption.[50] However, after noting that many jurisdictions had recently extended the rule to products intended for intimate bodily use, the ALI voted to correspondingly enlarge the scope of section 402A.[51] Yet the second draft of section 402A,[52] which was accepted at the 1962 ALI meeting,[53] still contained the provision that the seller would be subject to liability for "bodily harm."[54]

In 1964, although much of the Restatement (Second) of Torts was at the printer, Prosser resubmitted section 402A to the ALI due to the

---

46. Section 402A provides:

> (1) One who sells any product in a defective condition unreasonably dangerous to the user or consumer or to his property is subject to liability for physical harm thereby caused to the ultimate user or consumer, or to his property, if
>
> (a) the seller is engaged in the business of selling such a product, and
>
> (b) it is expected to and does reach the user or consumer without substantial change in the condition in which it is sold.
>
> (2) The rule stated in Subsection (1) applies although
>
> (a) the seller has exercised all possible care in the preparation and sale of his product, and
>
> (b) the user or consumer has not bought the product from or entered into any contractual relation with the seller.

Restatement (Second) of Torts § 402A (1977).

47. This is the title of § 402A. *See* Restatement, *supra* note 46, § 402A.

48. *See* 38 A.L.I. Proc. 49-58 (1961).

49. *See id.* at 64. "[A]s long as the section is limited to food, we are limiting it to personal injury—injury to the body." *Id.*

50. Restatement (Second) of Torts (Tent. Draft No. 6, at 24 (1961)). "One engaged in the business of selling food for human consumption who sells such food in a defective condition unreasonably dangerous to the consumer is subject to liability for *bodily harm* thereby caused to one who consumes it . . . ." *Id.* (Emphasis added).

51. *See* A.L.I. Proc., *supra* note 48, at 75. In addition, the movement toward strict liability for all products was recognized through the use caveats. *Id.* at 85-86.

52. Restatement (Second) of Torts (Tent. Draft No. 7, at 1 (1962)). "One engaged in the business of selling food for human consumption or other products for intimate bodily use, who sells such a product in a defective condition unreasonably dangerous to the consumer, is subject to liability for *bodily harm* thereby caused to one who consumes it . . . ." *Id.* (Emphasis added).

53. *See* 39 A.L.I. Proc. 244 (1962).

54. *See* Tent. Draft No. 7, *supra* note 52.

rapid development in the case law.[55] Prosser accordingly urged the ALI to expand section 402A to encompass all products.[56] With this revision, the term "physical harm" was substituted for "bodily harm."[57] Prosser explained that this semantic change was necessary to denote "not only physical injury, but also property damage; in other words, tangible damage."[58]

The ALI hurriedly passed the resubmitted version of section 402A,[59] and in the following decade a large majority of the jurisdictions adopted some form of strict tort liability against manufacturers and sellers for injuries caused by defective products.[60] Although subtle variations exist,

---

55. 40 A.L.I. Proc. 349 (1964) (both *Henningsen v. Bloomfield Motors, Inc.*, 32 N.J. 358, 161 A.2d 69 (1960) and *Greenman v. Yuba Power Products, Inc.*, 59 Cal. 2d 57, 377 P.2d 897, 27 Cal. Rptr. 697 (1963) had been decided by this time). In addition, a number of jurisdictions extended the strict liability to products for *external* bodily use, and several others were on the verge of such action. *Id.* at 350.

56. Prosser went so far as to predict that strict tort liability for injuries caused by defective products would be the majority rule within fifty years. *Id.*

57. Restatement (Second) of Torts (Tent. Draft No. 10, at 1 (1964). "One who sells any product in a defective condition unreasonably dangerous to the user or consumer or to his property, is subject to liability for *physical harm* thereby caused to the ultimate user or consmuer, or to his property . . . ." *Id.* (Emphasis added).

58. A.L.I. Proc., *supra* note 48, at 64. Prosser stated: "If you want to include other products, then I think you are going to have to include property damage, and make that 'physical harm.' " *Id.*

59. *See* A.L.I. Proc., *supra* note 55, at 375.

60. *See, e.g.*, Clary v. Fifth Ave. Chrysler Center, Inc., 454 P.2d 244 (Alaska 1969); O.S. Stapley Co. v. Miller, 103 Ariz. 556, 447 P.2d 248 (1968); Greenman v. Yuba Power Products, Inc., 59 Cal. 2d 57, 377 P.2d 897, 27 Cal. Rptr. 697 (1963); Bradford v. Bendix-Westinghouse Automotive Air Brake Co., 33 Colo. App. 99, 517 P.2d 406 (1973); Garthwait v. Burgio, 153 Conn. 284, 216 A.2d 189 (1965); Stewart v. Budget Rent-A-Car Corp., 52 Haw. 71, 470 P.2d 240 (1970); Shields v. Morton Chemical Co., 95 Idaho 674, 518 P.2d 857 (1974); Suvada v. White Motor Co., 32 Ill. 2d 612, 210 N.E.2d 182 (1965); Cornette v. Searjeant Metal Products, Inc., 147 Ind. App. 46, 258 N.E.2d 652 (1970); Hawkeye-Security Ins. Co. v. Ford Motor Co., 174 N.W.2d 672 (Iowa 1970); Dealers Transport Co. v. Battery Distrib. Co., 402 S.W.2d 441 (Ky. 1965); Spence v. Three Rivers Builders & Masonry Supply, Inc., 353 Mich. 120, 90 N.W.2d 873 (1958); Farr v. Armstrong Rubber Co., 288 Minn. 83, 179 N.W.2d 64 (1970); State Stove Mfg. Co. v. Hodges, 189 So. 2d 113 (Miss. 1966); Keener v. Dayton Elec. Mfg. Co., 445 S.W.2d 362 (Mo. 1969); Brandenburger v. Toyota Motor Sales, U.S.A., Inc., 162 Mont. 506, 513 P.2d 268 (1973); Kohler v. Ford Motor Co., 187 Neb. 428, 191 N.W.2d 601 (1971); Ginnis v. Mapes Hotel Corp., 86 Nev. 408, 470 P.2d 135 (1970); Buttrick v. Arthur Lessard & Sons, Inc., 110 N.H. 36, 260 A.2d 111 (1969); Glass v. Ford Motor Co., 123 N.J. Super. 599, 304 A.2d 562 (1973); Stang v. Hertz Corp., 83 N.M. 730, 497 P.2d 732 (1972); Codling v. Paglia, 32 N.Y.2d 330, 298 N.E.2d 622, 345 N.Y.S.2d 461 (1973); Johnson v. American Motors Corp., 225 N.W.2d 57 (N.D. 1974); Lonzrick v. Republic Steel Corp., 6 Ohio St. 2d 227, 218 N.E.2d 185 (1966); Kirland v. General Motors Corp., 521 P.2d 1353 (Okla. 1974); Heaton v. Ford Motor Co., 248 Or. 467, 435 P.2d 806 (1967); Webb v. Zenn, 422 Pa. 424, 220 A.2d 853 (1966); Ritter v. Narragansett

the strict tort liability generally adopted has been that denoted in section 402A or enunciated in *Greenman*.[61] The judiciary and the drafters of section 402A intended the manufacturer's strict tort liability to be a "special liability of sellers for physical harm to users or consumers."[62] Thus, a reasonable question is whether loss of consortium, being an intangible and nonpecuniary harm to one with a special relationship to the injured person, was contemplated as being within the scope of the "special" liability.

C. Society's Current Concern—Delimiting Liability

As previously noted, the strict tort liability concept evolved with an emphasis on consumer protection.[63] The courts generally reiterated the policies of loss-distribution or risk spreading, the availability of insurance, and injury reduction through enhanced safety.[64] As a result, litigants view section 402A as a panacea to correct all of the alleged wrongs inflicted by a product manufacturer and thereby tend to use the theory indiscriminately.[65] Yet the judiciary and the ALI never intended for section 402A to be an exclusive remedy for all product liability litigation[66]—negligence and breach of warranty theories were intended to retain their distinctive functions.[67] Additionally, established law holds that the man-

---

Elec. Co., 109 R.I. 176, 283 A.2d 255 (1971); Engberg v. Ford Motor Co., 87 S.D. 196, 205 N.W.2d 104 (1973); Ford Motor Co. v. London, 217 Tenn. 400, 398 S.W.2d 240 (1966); McKisson v. Sales Affiliates, Inc., 416 S.W.2d 787 (Tex. 1967); Zaleskie v. Joyce, 133 Vt. 150, 333 A.2d 110 (1975); Ulmer v. Ford Motor Co., 75 Wash. 2d 522, 452 P.2d 729 (1969); Dippel v. Sciano, 37 Wis. 2d 443, 155 N.W.2d 55 (1967). *See generally*, J. Beasley, *Products Liability and the Unreasonably Dangerous Requirement*, 101-339 (1981); Bieman, *Strict Products Liability: An Overview of State Law*, J. Prod. Liab., 111-178 (1987).

61. *See* Maleson, *Negligence Is Dead But Its Doctrines Rule Us From the Grave: A Proposal to Limit Defendant's Responsibility in Strict Products Liability Actions Without Resort to Proximate Cause*, 51 Temp. L.Q. 1, 38-40 (1978).

62. This is the title of § 402A. *See* Restatement, *supra* note 46, § 402A.

63. *See* Vandall, *supra* note 4, at 715; *see also* Hall v. E.I. Dupont de Nemours & Co., Inc., 345 F. Supp. 353, 368 (E.D.N.Y. 1972).

64. *See, e.g., Hall*, 345 F. Supp. at 368; *see also* Greenman v. Yuba Power Products, Inc., 59 Cal. 2d 57, 377 P.2d 897, 27 Cal. Rptr. 697 (1963).

65. *See* Manuel & Richards, *supra* note 4, at 315.

66. *See* Restatement, *supra* note 46, § 402A comment a.

67. Section 402A appears in Chapter 14 of the Restatement (Second) of Torts entitled "Liability of Persons Supplying Chattels for the Use of Others." This chapter also contains several sections imposing liability for *negligence* on suppliers of chattels—the basis of which is clearly fault. In addition, the continued role of the breach of warranty theory has been emphasized in cases such as *Seely v. White Motor Co.*, 63 Cal. 2d 9, 403 P.2d 145, 45 Cal. Rptr. 17 (1965), in which the court recognized the limited role intended for strict liability and noted that the remedy for purely economic or commercial loss remained in the law of sales.

ufactuer is not an insurer of the product and his strict liability may not be equated with absolute, limitless liability.[68]

However, until the mid-seventies, commentors generally applauded the expansion of the plaintiff's rights and remedies for product related injuries under strict tort liability.[69] Moreover, because products liability insurance was at that time relatively inexpensive and easy to obtain, there was little resistance from the business sector.[70] Thus, few cases, if any, presented a thorough and detailed analysis of whether the ever-expanding scope of the manufacturer's liability in fact promoted the perceived policy goals of *Greenman* and section 402A.[71]

Around 1975, however, concern about a "products liability crisis" surfaced.[72] In many jurisdictions, the complaint that the tort system had become a compensation system with tort damages led to significant movement toward tort reform.[73] The emphasis on consumer protection that was predominant in the sixties began to wane with the concern that the scales of justice were tilted too much in favor of the consumer.[74] Today, along with many other aspects of the tort and insurance systems, states are evaluating the effectiveness of damage awards in the product liability area and are attempting to delimit them through legislation.[75]

A similar judicial evaluation is now necessary to preserve the intended objectives of strict tort liability. Legitimate concerns exist about the overextension of the legal theory. Today, products liability insurance is not inexpensive and readily available.[76] Further, it is recognized that there is a limit to the range of injuries and the dollar amount of recovery which can be spread across society.[77] The manufacturer's strict tort

---

68. *See, e.g.*, Elliott v. Lachance, 109 N.H. 481, 256 A.2d 153, 156 (1969); Helene Curtis Industries, Inc. v. Pruitt, 385 F.2d 841, 849 (5th Cir. 1967); *see also* Traynor, *The Ways and Means of Defective Products and Strict Liability*, 32 Tenn. L. Rev. 363, 366-67 (1965).

69. *See* Reed & Watkins, *Product Liability Tort Reform: The Case For Federal Action*, 63 Neb. L. Rev. 389, 392 (1984); *see generally* Epstein, *Products Liability: The Search For the Middleground*, 56 N.C.L. Rev. 643 (1978); Prosser, *Assault Upon the Citadel*, *supra* note 4.

70. Reed & Watkins, *supra* note 69, at 392.

71. Shepard v. Superior Court, 76 Cal. App. 3d 16, ____, 142 Cal. Rptr. 612, 619 (1977) (Kane, J., dissenting).

72. Reed & Watkins, *supra* note 69, at 392.

73. *See* Dodd, *Proposal for Making Products Liability Fair, Efficient, and Predictable*, 14 J. of Legis. 133 (1987).

74. *See generally* Vandall, *supra* note 4.

75. *See generally* Gilmartin, *States and Trends: Punitive Damages*, 14 J. of Legis. 249 (1987); Dodd, *supra* note 73.

76. *See generally* Reed & Watkins, *supra* note 69.

77. Berger v. Weber, 411 Mich. 1, ____, 303 N.W.2d 424, 438 (1981) (Levin, J., dissenting).

liability must be carefully circumscribed if it is to be preserved as the remedy intended by Judge Traynor and the ALI for the primary victims of the multitude of today's product related injuries.

In delimiting the scope of a tortfeasor's liability, a court "must take into account considerations in addition to logical symmetry and sympathetic appeal;"[78] a court must balance the need for compensation against public policy considerations and social consequences.[79] "[T]he law cannot redress every injury, and the determination of where to draw the line of liability is essentially a question of policy."[80] Thus, considering today's emphasis on the relational, intangible and nonpecuniary aspects of loss of consortium,[81] a court could reasonably determine that there are now sufficient countervailing policies and social ramifications present to preclude the cause of action within a section 402A products liability case.

III. Analysis of the Arguments

A. Overcoming Precedent

Because valid arguments support both sides of this issue, an initial factor to address is that of precedent. Courts have consistently allowed the loss of consortium claim in products liability actions, almost without question.[82] Thus, one could logically assert that recovery should continue. Yet determining the extent of tort liability is more than an exercise in logic.[83] It is a pronouncement of social policy which should reflect the subtle balance of the interests involved.[84]

---

78. Borer v. American Airlines, Inc., 19 Cal. 3d 441, ____, 563 P.2d 858, ____, 138 Cal. Rptr. 302, 306 (1977).

79. Sizemore v. Smock, 430 Mich. 283, ____, 422 N.W.2d 666, 670 (1988).

80. *Id.* at ____, 422 N.W.2d at 671.

81. *See supra* notes 21-24 and accompanying text.

82. *Cf.* Park v. Standard Chem. Way Co., 60 Cal. App. 3d 47, 131 Cal. Rptr. 338, 340 (1976) (Compton, J., concurring). This is the only case in which it was noted, in the concurring opinion, that the loss of consortium claim should not be compensable in an action based on a manufacturer's or retailer's strict liability. Judge Compton relied upon the reasoning of *Seely v. White Motor Co.*, 63 Cal. 2d 9, 403 P.2d 145, 45 Cal. Rptr. 17 (1965)—that the manufacturer's liability should not be of unknown and unlimited scope.

83. Berger v. Weber, 411 Mich. 1, ____, 303 N.W.2d 424, 430 (1981) (Levin, J., dissenting). *See also*, Shepard v. Superior Court, 76 Cal. App. 3d 16, ____, 142 Cal. Rptr. 612, 621 (1977) (Kane, J., dissenting); Borer v. American Airlines, Inc., 19 Cal. 3d 441, ____, 563 P.2d 858, ____, 138 Cal. Rptr. 302, 305 (1977).

84. *Berger*, 411 Mich at ____, 303 N.W.2d at 430.

Both the manufacturer's strict liability in tort and the loss of consortium claims are essentially judicial creations,[85] and are thus subject to the scrutiny of traditional common law. The nature of the commom law requires that each time a rule of law is applied, the courts conduct a deliberate examination of the underlying policies to ensure that the conditions and needs of the times have not changed so as to make further application of the rule an instrument of injustice.[86] The inherent capacity of the common law for growth and change is a most significant feature. "But that vitality can flourish only so long as the courts remain alert to their obligation and opportunity to change the common law when reason and equity demand it . . . ."[87] There is then, a judicial duty to evaluate the loss of consortium claim to determine if equity and reason now call for a change in the law.

Further, although it could be asserted that this type of change would be better left to the legislature, the courts have frequently exercised their power to change the law within the fields of products liability and loss of consortium. As one court stated: "We find no wisdom in abdicating to the legislature our essential function of re-evaluating common-law concepts in the light of present day realities."[88] In additon, judicial action in no way precludes a legislative appraisal of the issue: the legislature is still free to act by way of ratification, limitation, or rejection of judicial holdings.[89] Therefore, because judicial analysis of whether the loss of consortium cause of action should remain within the scope of strict tort products liability is not constrained by precedent, the following arguments merit consideration.

B. Section 693 of the Restatement (Second) of Torts

Although the original Restatement of Torts recognized the husband's loss of consortium claim, it was not until 1969 that the American Law Institute adopted the wife's cause of action for loss of consortium.[90] Section 693 of the Restatement (Second) of Torts provides that one is

---

85. *See* Weitl v. Moes, 311 N.W.2d 259, 266 (Iowa 1981) (loss of consortium is a creation of the common law); Reed & Watkins, *supra* note 69, at 391 (products liability law is essentially judge made).

86. 15A Am. Jur. 2d, *Commom Law*, § 3, 599 (1976).

87. Rodriguez v. Bethlehem Steel Corp., 12 Cal. 3d 382, ____, 525 P.2d 669, ____, 115 Cal. Rptr. 765, 772 (1974).

88. Dini v. Naiditch, 20 Ill. 2d 406, ____, 170 N.E.2d 881, 892 (1960). *See also* Landes & Posner, *A Positive Economic Analysis of Products Liability*, 14 J. Legal Stud. 535, 551 (1985) (it can be implied that legislatures act in common-law fields such as torts only after the judiciary has failed to alter existing doctrine).

89. Hay v. Medical Center Hosp., 145 Vt. 533, ____, 496 A.2d 939, 946 (1985).

90. *See* 46 A.L.I. Proc. 148-157 (1969).

subject to liability for loss of consortium if "by reason of his tortious conduct [he] is liable to [the other] spouse for illness or other bodily harm . . . ."[91] Yet a manufacturer's strict tort liability is not predicated upon "tortious" or wrongful conduct.[92] The touchstone of section 402A is the presence of a defective product placed into the stream of commerce.[93] The manufacturer's conduct is not a determinative factor; rather, the liability is based largely upon a policy determination that the manufacturer or seller is in a better position to absorb the loss than the consumer.[94] Thus, a reasonable interpretation of section 693 is that strict products liability would not render the manufacturer liable for loss of consortium.

However, the commentary to section 693 notes that the word "tortious" includes "conduct intended to cause harm, conduct that is negligent and conduct that is carried on at the risk of the actor."[95] The original Restatement, as well as the tentative draft of this section presented to the ALI at its 1969 meeting, contained the above comment.[96] The question then becomes whether manufacturing can be deemed conduct carried on at the risk of the actor. The original Restatement and the tenative draft considered by the ALI further explained the category of "conduct carried on at the risk of the actor." Significantly, the final version of this commentary contained a portion not presented to nor debated by the ALI at its 1969 meeting:

> Thus one who has become liable to a spouse because of conducting an abnormally dangerous activity or because of harm inflicted by a wild animal in his possession or *because of supplying a defective and unreasonably dangerous product*, is subject to liability to the deprived spouse under the rule stated in this

---

91. Section 693 provides:

> (1) One who by reason of his tortious conduct is liable to one spouse for illness or other bodily harm is subject to liability to the other spouse for the resulting loss of the society and services of the first spouse, including impairment of capacity for sexual intercourse, and for reasonable expense incurred by the second spouse in providing medical treatment.
>
> (2) Unless it is not possible to do so, the action for loss of society and services is required to be joined with the action for illness or bodily harm, and recovery for loss of society and services is allowed only if the two actions are so joined.

RESTATEMENT (SECOND) OF TORTS § 693 (1977).

92. *See, e.g.*, Shepard v. Superior Court, 76 Cal. App.3d 16, ____, 142 Cal. Rptr. 612, 618 (1977) (Kane, J., dissenting) (the manufacturer's negligence or culpability is not a necessary ingredient in strict products liability actions).

93. *See* RESTATEMENT, *supra* note 46, § 402A comment j.

94. *See supra* notes 36-45 and accompanying text.

95. *See* RESTATEMENT, *supra* note 91, § 693 comment b.

96. *See* RESTATEMENT (SECOND) OF TORTS (Tent. Draft No. 14, at 5 (1969)).

Section, as well as one who has become liable because of an intentional or negligent injury to the impaired spouse.[97]

The tentative draft considered by the ALI did not include the underlined clause relating to strict products liability.[98] The Reporter, Dean Prosser, therefore made this inclusion largely on his own. Yet there was very little case law that Prosser could have used as justification for this addition—even the Reporter's Note indicates that the vast majority of cases involving loss of consortium were based on negligence.[99]

Further, a limitation found implicitly in the relevant case law should be considered. Decisions which have adopted loss of consortium as a cause of action express the claim in terms of a negligent or intentional infringement of a legally cognizable right. For example, in *Rodriguez v. Bethlehem Steel Corp.*,[100] the Supreme Court of California adopted the wife's loss of consortium claim and based its holding upon a recognition of liability "as it is currently understood by the large preponderance of our sister states and a consensus of distinguished legal scholars."[101] The court then declared that "each spouse has a cause of action for loss of consortium . . . caused by a negligent or intentional injury to the other spouse by a third party."[102] Although a few recent courts patterned their holdings after the Restatement (Second) language,[103] most decisions have continued to limit holdings in a manner similar to that of *Rodriguez*.[104] Moreover, the very recent cases recognizing loss of filial consortium have holdings limiting the liability to injuries resulting from negligence.[105]

Additionally, it is widely recognized that Prosser exaggerated the supporting case law in his zeal to establish the manufacturer's strict tort liability denoted in section 402A.[106] Thus, Prosser's inclusion of the

---

97. *See id.* (emphasis added).

98. *See* A.L.I. Proc., *supra* note 90, at 148-157; Tent. Draft No. 14, *supra* note 96, at 5.

99. *See* Restatement, *supra* note 91, § 693 (Appendix) at 514-15.

100. 12 Cal. 3d 382, 525 P.2d 669, 115 Cal. Rptr. 765 (1974).

101. *Id.* at ____, 525 P.2d at ____, 115 Cal. Rptr. at 771.

102. *Id.* at ____, 525 P.2d at ____, 115 Cal. Rptr. at 782.

103. A few courts used the "tortious conduct" language. *See, e.g.*, Ueland v. Reynolds Metals Co., 103 Wash. 2d 131, 691 P.2d 190 (1984); Hay v. Medical Center Hosp., 145 Vt. 533, 496 A.2d 939 (1985).

104. *See, e.g.*, Diaz v. Eli Lilly and Co., 302 N.E.2d 555 (Mass. 1973); Ekalo v. Constructive Serv. Corp., 46 N.J. 82, 215 A.2d 1 (1965); Millington v. Southeastern Elevator Co., 22 N.Y.2d 498, 239 N.E.2d 897, 293 N.Y.S.2d 305 (1968); Clouston v. Remlinger Oldsmobile Cadillac, Inc., 22 Ohio St. 2d 65, 258 N.E.2d 230 (1970).

105. *See, e.g.*, Ferriter v. Daniel O'Connell's Sons, 381 Mass. 507, 413 N.E.2d 690 (1980); Berger v. Weber, 411 Mich. 1, 303 N.W.2d 424 (1981).

106. *See* Priest, *The Invention of Enterprise Liability: A Critical History of the*

above clause in the commentary to section 693 can legitimately be questioned as a "restatement of the existing case law" regarding loss of consortium under a strict tort liability theory. As a result, the issue becomes whether the manufacturer's strict liability in tort *should* be categorized with the other strict liability denoted in the commentary—that resulting from abnormally dangerous activities or the keeping of a wild animal.

Strict liability resulting from the keeping of a wild animal evolved out of the primitive law which held the owner of property strictly liable for the harm it did.[107] The survival of the primitive notion has been due to modern policy considerations.[108] Certain animals involve an obvious danger to the community; thus, those who keep such animals for their own purposes are required to protect the community, at their peril, against the risk involved.[109]

The roots of strict liability for the conducting of an abnormally dangerous activity extend to *Rylands v. Fletcher*,[110] which imposed an absolute duty on one who collected on his land something non-natural and likely to do mischief if it escaped.[111] Early American blasting cases[112] imposed liability without fault largely because the defendant elected to engage in this perilous activity as a business.[113] The underlying rationale for this type of strict liability was that even if the defendant exercised reasonable care, there was still a high degree of risk of harm,[114] as compared with negligence in which it is presumed that reasonable care will decrease the risk of harm.

In contrast, the rationales underlying strict products liability are readily distinguishable. First, if the manufacturer exercises due care, there

---

Intellectual Foundations of Modern Tort Law, 14 J. LEGAL STUD. 461, 514-17 (1985) (today it is widely recognized that Prosser greatly exaggerated the case law supporting 402A). *See generally* Prosser, *Fall of the Citadel*, *supra* note 3; Prosser, *Assault Upon the Citadel*, *supra* note 3.

107. DOBBS, KEETON & OWEN, PROSSER & KEETON ON TORTS, § 76, at 538 (5th Ed. 1984).

108. *Id.*

109. *Id.* Consider for example, the propensity of cattle and horses to escape and roam and do mischief. *See, e.g.*, Page v. Hollingsworth, 7 Ind. 317 (1855); Gresham v. Taylor, 51 Ala. 505 (1874).

110. Fletcher v. Rylands, 3 H. & C. 774, 159 Eng. Rep. 737 (1865), *rev'd*, L.R. 1 Ex. 265 (1866), *aff'd*, Rylands v. Fletcher, L.R. 3 H.L. 332 (1868).

111. *Id.* If the thing likely to do mischief escaped, the owner was prima facie answerable for all damages which were the natural consequence of its escape. However, assumption of the risk and an "act of God" would constitute valid defenses. *Id.*

112. *See, e.g.*, Blackford v. Heman Construction Co., 132 Mo. App. 157, 112 S.W. 287 (1908); Dixon v. New York Trap Rock Corp., 293 N.Y. 509, 58 N.E.2d 517 (1944); Gossett v. Southern Ry. Co., 115 Tenn. 376, 89 S.W. 737 (1905).

113. *Id.*

114. *See* DOBBS, KEETON, & OWEN, *supra* note 107, § 78, at 554-56.

is not going to be a high degree of risk of harm. Defective products are generally the exception, not the norm.[115] Second, a predominant justification of strict tort liability is that it will enhance product safety;[116] this implies that the level of care exercised by a manufacturer will affect the degree of the risk of harm.[117] Third, rather than placing the loss on the defendant because of his election to conduct the business, the liability is imposed as a means of shifting the loss to the manufacturer because he is in a better position to absorb and distribute that loss among all consumers.[118]

Hence, the strict tort liability imposed on a manufacturer does not inherently belong within the same category of the strict liability enunciated in the original Restatement provision dealing with loss of consortium. Indeed, placing products liability within the same class as abnormally dangerous activities could arguably tend to discourage manufacturing and thereby adversely affect our national economy—in which case public policy would preclude such a classification. Therefore, although section 693 of the Restatement (Second) of Torts provides that it is appropriate to subject the manufacturer to strict liability for loss of consortium, it is lacking both the proper foundation and the thorough analysis necessary to justify such a pronouncement.

C. Section 402A of the Restatement (Second) of Torts

1. Express Limitations.—An analysis of valid recovery under strict tort products liability must also focus on the specific language of section 402A of the Restatement (Second) of Torts.[119] Section 402A expressly limits the manufacturer's liability to "physical harm thereby caused to the ultimate user or consumer, or to his property . . . ."[120] Dean Prosser

---

115. An essential requirement for a § 402A action is that a "defect" must cause the harm. "[A] plaintiff must trace his injury to a quality or condition of the product which was unreasonably dangerous either for a use to which the product would ordinarily be put, or for some special use which was brought to the attention of the defendant." James, *supra* note 40, at 927.

116. *See, e.g.*, Hall v. E.I. DuPont de Nemours & Co., Inc., 345 F. Supp. 353, 368 (E.D.N.Y. 1972) (a rigorous rule of liability with enhanced possibilities of large recoveries is an "incentive" to maximize safe design).

117. *Id.* "A manufacturer is in the best position to discover defects or dangers in his product and to guard against them through appropriate design, manufacturing and distribution safeguards, inspections and warnings." *Id.*

118. *Id.* "Regardless of safety measures taken by manufacturer's and distributors, accidents and injuries will inevitably occur. . . . Accidents and injuries, in this view, are seen as an inevitable and statistically foreseeable 'cost' of the product's consumption or use." *Id.*

119. *See* RESTATEMENT, *supra* note 46, § 402A.

120. Id.

indicated that the deliberate use of "physical harm" was necessary if the section was to apply to all products in order to limit the liability to "tangible harms."[121] Comment d of section 402A states that the rule obviously includes the sale of food for human consumption and other products intended for intimate bodily use.[122] Prosser noted during the ALI debates that if section 402A had been limited to these products, the rule would have been drafted with the use of "bodily harm;"[123] because a defective food product or product intended for intimate bodily use would naturally result in harm to the body.[124]

However, comment d further states that the section applies "also to products which, if they are defective, may be expected to and do cause only 'physical harm' in the form of damage to the user's land or chattels . . . ."[125] This language appears to contemplate liability for either (1) the type of "bodily harm" to a person which would naturally be caused by a defective food product or product intended for intimate bodily use, or (2) "physical harm" in the form of damage to the user's land or chattels caused by any defective product.

Furthermore, one year after the adoption of section 402A by the ALI, Professor Keeton indicated that there were four categories of harm that could result from the use of a defective product: (1) physical injury to persons; (2) physical damage to tangible things other than the product; (3) physical harm to the product itself; and (4) commercial or economic loss.[126] Thus, nebulous and intangible harms, such as emotional distress and loss of consortium, were not contemplated as being within the scope of section 402A when drafted.

Section 402A further limits the manufacturer's liability to "users or consumers" of the defective product.[127] Interestingly, in spite of these

---

121. *See* A.L.I. Proc., *supra* note 48, at 64.

122. Restatement, *supra* note 46, § 402A comment d.

123. *See* A.L.I. Proc., *supra* note 48, at 64.

124. *Id.* at 55.

125. Restatement, *supra* note 46, § 402A comment d.

126. *See* Keeton, *Products Liability—Some Observations About Allocations of Risks*, 64 Mich. L. Rev. 1329, 1343-44 (1966). However, most courts do not allow recovery for economic loss; *see also* Manuel & Richards, *supra* note 4, at 321.

127. The Restatement explains the terms "user or consumer" as follows: " 'Consumers' include not only those who in fact consume the product, but also those who prepare it for consumption; and a housewife who contracts tularemia while cooking rabbits for her husband is included within the rule stated in this Section, as is also the husband who is opening a bottle of beer for his wife to drink. Consumption includes all ultimate uses for which the product is intended, and the customer in a beauty shop to whose hair a permanent wave solution is applied by the shop is a consumer. 'User' includes those who are passively enjoying the benefit of the product, as in the case of passengers in automobiles or airplanes, as well as those who are utilizing it for the purposes of doing

express limitations, the courts have reached inconsistent results when a user or consumer has sought recovery for *emotional harms* in strict products liability cases.[128] In fact, only relatively recently have courts recognized a cause of action for the *negligent* infliction of emotional distress; and some jurisdictions still stringently limit such a claim.[129] In *Rahn v. Gerdts*,[130] the Illinois Appellate Court retained a restrictive interpretation of the "physical harm" requirement of section 402A and held that the plaintiff's depression, anxiety and nervousness caused by emotional distress, unaccompanied by other physical injury, were not compensable in a strict tort liability action.[131]

In contrast, in *Walters v. Mintec International*,[132] the Third Circuit Court of Appeals found the evidence of plaintiff's headaches, weakness under stress, insomnia and nightmares sufficient to satisfy the "physical harm" requirement.[133] This inconsistency, when the plaintiff seeking redress for emotional harm is the actual user or consumer of the defective product, implies an even greater need for analysis and justification for recovery of emotional distress or loss of consortium in cases in which the plaintiff is not the user or consumer. Thus, because the plaintiff in a suit for loss of consortium is generally not the user or consumer of the defective product, the courts should not automatically award such damages without an appropriate analysis of the arguments both for and against recovery.

---

work upon it, as in the case of an employee of the ultimate buyer who is making repairs upon the automobile which he has purchased." RESTATEMENT, *supra* note 46, § 402A comment l; *see, e.g.*, Cottom v. McGuire Funeral Serv., Inc., 262 A.2d 807, 809 (D.C. 1970) (person utilizing product for the purpose for which it was made is a user of that product).

128. There have been many cases dealing with recovery for psychological reactions to the discovery of foreign objects or substances in food or drink; in these cases the courts generally allow recovery. *See, e.g.*, Kroger Co. v. Beck, 176 Ind. App. 202, 204-05, 375 N.E.2d 640, 643 (1978) (hyperdermic needle in steak); Shoshone Coca-Cola Bottling Co. v. Dolinski, 82 Nev. 439, 441-42, 420 P.2d 855, 857 (1966) (decomposed mouse in bottled beverage). *See generally* R. DICKERSON, PRODUCTS LIABILITY AND THE FOOD CONSUMER, 236-37 (1951); Mead, *Recovery for Psychic Harm in Strict Products Liability: Has the Interest in Psychic Equilibrium Come the Final Mile?*, 35 DEF. L.J. 193, 242-43 (1986).

129. As of 1985, many jurisdictions still required physical impact for recovery of emotional distress. *See, e.g.*, Rickey v. Chicago Transit Authority, 98 Ill. 2d 546, 551, 457 N.E.2d 1, 5 (1983); Dziokonski v. Babineau, 375 Mass. 555, 559, 380 N.E.2d 1295, 1297-98 (1978); Bass v. Nooney Co., 646 S.W.2d 765, 771-72 (Mo. 1983) (en banc); Battalla v. State, 10 N.Y.2d 237, 240, 176 N.E.2d 729, 730, 219 N.Y.S.2d 34, 35-36 (1961).

130. 119 Ill. App. 3d 781, 455 N.E.2d 807 (1983).

131. *Id.* at 784-85, 455 N.E.2d at 809.

132. 758 F.2d 73 (3d Cir. 1985).

133. *Id.* at 78-79.

2. *Bystander Analogy.*—Significantly, the plaintiff in a loss of consortium action is generally not a "user or consumer."[134] Of course, all jurisdictions extended section 402A to include physically injured bystanders.[135] Yet a strict products liability "bystander" is a "non-user" of the product. Thus, allowing bystander recovery is in conflict with the black letter of section 402A.[136] The courts therefore made a policy determination that a bystander deserves the same protection as the consumer. In *Elmore v. American Motors Corp.*,[137] the court noted that privity had no place in strict products liability and that bystanders should recover for physical injuries.[138] The court further noted that a bystander might need even more protection than the user since he generally has no knowledge of the product or its warnings.[139]

Accordingly, a relevant issue in an analysis of whether loss of consortium is a valid recovery under strict products liability is whether the loss of consortium plaintiff is sufficiently akin to a bystander to reap the benefits of bystander status. Generally, a requisite for classification as a bystander is close proximity to the product user. The bystander is either physically injured in some way by the product or, in the case of an emotionally injured bystander, is an actual witness to the user's tragic injury. In contrast, such a requirement is not essential to the loss of consortium plaintiff. Further, in negligence cases, courts have adopted various tests to determine whether a bystander could recover for emotional distress.

For example, in the leading case of *Dillon v. Legg*,[140] recovery depended on whether the plaintiff (1) was in physical proximity to the scene of the accident, (2) had a contemporaneous sensory perception of the accident, and (3) had a close familial relationship with the physically injured party.[141] Yet the loss of consortium plaintiff can generally satisfy

---

134. *See supra* note 127.

135. Bystander recovery was first permitted in *Elmore v. American Motors Corp.*, 70 Cal. 2d 578, 451 P.2d 84, 75 Cal. Rptr. 652 (1969). The court in *Elmore* noted that privity has no place in strict products liability and bystanders may even need more protection than the user or consumer. *See also* Caruth v. Mariani, 11 Ariz. App. 188, 463 P.2d 83 (1970); Mitchell v. Miller, 26 Conn. Supp. 142, 214 A.2d 694 (1965); Sills v. Massey - Ferguson, Inc., 296 F. Supp. 776 (N.D. Ind. 1969); Lamendola v. Mizell, 115 N.J. Super. 514, 280 A.2d 241 (1971); Darryl v. Ford Motor Co., 440 S.W.2d 630 (Tex. 1969).

136. A caveat to § 402A states: "The Institute expresses no opinion as to whether the rules stated in this Section may not apply (1) to harm to persons other than users or consumers . . . ." RESTATEMENT, *supra* note 46, § 402A caveat (1).

137. 70 Cal. 2d 578, 451 P.2d 84, 75 Cal. Rptr. 652 (1969).

138. *Id.* at ____, 451 P.2d at 88-89, 75 Cal. Rptr. at 656.

139. *Id.* at ____, 451 P.2d at 89, 75 Cal. Rptr. at 657.

140. 68 Cal. 2d 728, 441 P.2d 912, 69 Cal. Rptr. 72 (1968).

141. *Id.* at 739-41, 441 P.2d at 919-21, 69 Cal. Rptr. at 79-81.

only the requirement of a close familial relationship. Thus, it is reasonable to assert that the loss of consortium plaintiff is not sufficiently similar to the bystander to gain analogous treatment. However, even if that status were to be conceded, loss of consortium recovery under section 402A is still debatable because the courts have experienced difficulty extending the rationales permitting recovery for the physically injured bystander in strict products liability cases to the emotionally injured bystander.

California was the first jurisdiction to consider whether a bystander suffering emotional harm as a result of an injury to a third party could recover under a strict products liability theory.[142] In *Park v. Standard Chemical Way Co.*,[143] the plainfiff's husband was injured when a container of drain cleaner exploded causing severe burns on his arms, torso and legs. The plaintiff, who returned from work two and a half hours after the explosion and found her husband permanently scarred, alleged severe emotional distress and anxiety.[144] The court recognized the husband's action for his injuries based on a theory of strict tort liability,[145] but rejected the wife's claim for emotional distress, also based on strict liability.[146] The court noted that *Dillon v. Legg* did not apply because the plaintiff did not allege the defendant's negligence and because she was not within the zone of danger at the time of the accident.[147]

However, the Court of Appeals for the First District of California ruled otherwise in *Shepard v. Superior Court*.[148] In *Shepard*, a family was riding in its automobile. Another car hit the vehicle, the rear door opened due to a defective locking mechanism, and the daughter fell out and was killed when struck by an oncoming car. The parents and brother brought a strict products liability action against the manufacturer of the car to recover for their emotional harm.[149] This court applied the *Dillon* test, labelled the plaintiffs Dillon-type bystanders and allowed recovery for their emotional distress in a strict products liability case.[150] The court noted that it would make no sense to permit recovery against a negligent driver, as in *Dillon*, while denying recovery from a manufacturer responsible for a defective product.[151] The court thus based its decision upon a policy rationale.

---

142. *See* Park v. Standard Chem Way Co., 60 Cal. App. 3d 47, 50, 131 Cal. Rptr. 338, 339 (1976); *see also* Mead, *supra* note 128, at 248-259.

143. 60 Cal. App. 3d 47, 131 Cal. Rptr. 338 (1976).

144. *Id.* at 49, 131 Cal. Rptr. at 339.

145. *Id.* at 50, 131 Cal. Rptr. at 339.

146. *Id.*

147. *Id.*

148. 76 Cal. App. 3d 16, 142 Cal. Rptr. 612 (1977).

149. *Id.* at 18, 142 Cal. Rptr. at 613.

150. *Id.* at ____, 142 Cal. Rptr. at 615.

151. *Id.*

However, the court in *Shepard* did not allude to the Restatement's limitation to physical harm. Further, the *Shepard* dissent emphasized that in *Dillon* there was a focus on the defendant's fault, and aptly noted that the policies underlying strict products liability, in which the manufacturer's culpability is not a necessary ingredient, would seem to preclude recovery.[152]

California again faced this issue in *Kately v. Wilkinson*.[153] In *Kately*, the plaintiff owner of a speed boat struck and killed a family friend who was water skiing because a defect in the boat caused the steering mechanism to malfunction. The court in *Kately* relied on *Shepard* to find that a plaintiff could recover for emotional distress in a products liability case.[154] However, this plaintiff could not recover under the *Dillon* test, and the court therefore utilized a test based upon the directness of the injury.[155] The court wanted to avoid bystander limitations and permit recovery for policy reasons.

New York has also considered recovery for emotional harm based on strict products liability. In *Vaccaro v. Squibb Corp.*,[156] the plaintiffs were parents of a child born with severe defects allegedly caused by an anti-miscarriage drug administered to the mother during her pregnancy. The Appellate Division, distinguishing between the mother as a user of the drug and the father as a bystander, allowed the mother's claim but dismissed the father's because bystanders may not recover for emotional harm in New York.[157] However, the Court of Appeals, focusing on the mother as a bystander suffering through concern for her deformed child, dismissed the mother's cause of action.[158]

The Illinois Court of Appeals also expressly stated that bystander recovery for emotional distress was not compensable in strict products liability actions in *Woodhill v. Parke Davis & Co.*[159] In *Woodhill*, the

---

152. *Id.* at ____, 142 Cal. Rptr. at 618 (Kane, J., dissenting).

153. 148 Cal. App. 3d 576, 195 Cal. Rptr. 902 (1983).

154. *Id.* at ____, 195 Cal. Rptr. at 908.

155. *Id.* The court in *Kately* used a test developed in *Molien v. Kaiser Found. Hosps.*, 27 Cal. 3d 916, 616 P.2d 813, 167 Cal. Rptr. 831 (1980) (if the injury was direct, the three prong test of *Dillon* was not controlling).

156. 71 A.D. 2d 270, 422 N.Y.S.2d 679 (1979), *rev'd*, 52 N.Y.2d 809, 418 N.E.2d 386, 436 N.Y.S.2d 871 (1980).

157. *Id.* at 277-78, 422 N.Y.S.2d at 683-84. (*Tobin v. Grossman*, 24 N.Y.2d 609, 611, 249 N.E.2d 419, 419-20, 301 N.Y.S.2d 554, 555 (1969), had previously established that a bystander may not recover for emotional harm caused by witnessing an accident that causes severe physical injury to a third person).

158. *Vaccaro*, 52 N.Y.2d at 812, 418 N.E.2d at 387, 436 N.Y.S.2d at 872 (1980) (Fuchsberg, J., dissenting).

159. 58 Ill. App. 3d 349, 374 N.E.2d 683 (1978), *aff'd*, 79 Ill. 2d 26, 402 N.E.2d 194 (1980).

mother of a child born with brain damage claimed emotional harm allegedly caused by the administration of pitocin, a drug used to induce labor. The court dismissed the action and, in dictum, stated that because section 402A expressly limits recovery to cases involving "physical harm," the Restatement did not intend to provide a strict liability action for mental anguish or emotional harm.[160]

Thus, even if the loss of consortium plaintiff could be labeled as an emotionally injured bystander, this alone would not validate recovery under section 402A. Further, the fact that the courts are struggling with the issue of recovery for an emotionally injured bystander is significant. Bystander recovery is rendered dependent on various tests and limitations to appropriately delineate the extent of a tortfeasor's responsibility to a plaintiff other than the primary victim. Yet a loss of consortium plaintiff is generally even more remote than a bystander,[161] and liability should thus be similarly circumscribed. However, the courts in general have failed to conduct the analysis of loss of consortium necessary for a proper delimitation of the manufacturer's responsibility under a strict tort liability theory.

3. *Examples of Inadequate Judicial Analysis.*—As previously noted, the Illinois courts do not permit recovery for emotionally injured bystanders under a strict products liability theory.[162] Yet recovery for loss of consortium in section 402A actions is permitted.[163] In *Hammond v. North American Asbestos Corp.*,[164] the wife of an asbestos worker who contracted asbestosis brought a strict products liability action against a manufacturer for loss of consortium.[165] The loss of consortium claim was dependent upon a "tortious" injury and the court thus conducted a lengthy analysis of whether the husband's injury properly met the requirements of a strict tort liability cause of action.[166] However, once that initial determination was affirmatively answered, the court only briefly questioned the validity of the wife's claim for loss of consortium in light of a statute of limitations problem.[167] Thus, the court presumed that loss of consortium was valid under section 402A, notwithstanding its prior holding in *Woodhill v. Parke Davis & Co.*[168] which expressly

160. *Id.* at ____, 374 N.E.2d at 688.

161. *See infra* notes 179-83 and accompanying text.

162. *See supra* notes 159-60 and accompanying text.

163. *See* Hammond v. North Am. Asbestos Corp., 97 Ill. 2d 195, 454 N.E.2d 210 (1983).

164. *Id.*

165. *Id.* at ____, 454 N.E.2d at 213.

166. *Id.* at ____, 454 N.E.2d at 215-18.

167. *Id.* at ____, 454 N.E.2d at 218.

168. 58 Ill. App. 3d 349, 374 N.E.2d 683 (1978), *aff'd*, 79 Ill. 2d 26, 402 N.E.2d 194 (1980). *See supra* notes 159-60 and accompanying text.

limited recovery under section 402A to "physical harm." Interestingly, an Illinois court very recently denied a claim for loss of filial consortium in a strict products liability case in which the concurring opinion, noting that *Woodhill* should be controlling, emphasized that section 402A limits recovery to "physical harm."[169]

Also, in *Park v. Standard Chemical Co.*,[170] discussed above as denying recovery to an emotionally injured bystander in strict products liability actions,[171] the plaintiff alleged a "partial loss of consortium" cause of action.[172] The court denied this claim merely because recovery for loss of consortium required a complete loss for a definite period of time, or for a non-determinable period of time.[173] Therefore, although the court did not permit recovery as an emotionally injured bystander, it would have allowed the loss of consortium cause of action had it been properly pled. The concurring opinion would have gone further and denied the loss of consortium claim as noncompensable within strict products liability.[174] Noting that the manufacturer's liability is premised upon a social policy, the concurring opinion stated that the manufacturer should not be subject to damages of unknown and unlimited scope.[175] This was the only opinion the author discovered which even questioned the validity of loss of consortium under a theory of strict tort liability.

The inconsistencies and arbitrary delimiting devices highlighted by the foregoing cases indicate the complex considerations the judiciary faces when evaluating claims for intangible damages in strict products liability cases. There must be a proper balance between the need for adequate recovery and the principle that the manufacturer is not an insurer of the product.[176] Because the judiciary must strive to maintain logical limits on a tortfeasor's responsibility, the relational loss of consortium claim within the context of liability without fault should be the subject of thorough analysis.

D. Economic Considerations

As a matter of judicial policy, the courts could reasonably preclude the extension of the manufacturer's strict liability to areas where, as

---

169. *See* Dralle v. Ruder, 124 Ill. 2d 61, 529 N.E.2d 209 (1988).

170. 60 Cal. App. 3d 47, 131 Cal. Rptr. 338 (1976).

171. *See supra* notes 143-47 and accompanying text.

172. *Park*, 60 Cal. App. 3d at ____, 131 Cal. Rptr. at 339 (allegation 9).

173. *Id.* at ____, 131 Cal. Rptr. at 340 (loss of consortium was not to be confused with the inevitable physical, mental and emotional damage normally suffered by one spouse when the other has been wrongfully injured).

174. *Id.* at 51, 131 Cal. Rptr. at 340 (Compton, J., concurring).

175. *Id.*

176. Shepard v. Superior Court, 76 Cal. App. 3d 16, 29, 142 Cal. Rptr. 612, 620 (1977) (Kane, J., dissenting); *see also* Helene Curtis Industries, Inc. v. Pruitt, 385 F.2d 841, 862 (5th Cir. 1967).

here, the injury is to a relationship and the damages are intangible and speculative. "The loss of companionship, emotional support, love, felicity, and sexual relations are real injuries."[177] Yet the liability is secondary.[178] The loss of consortium plaintiff does not bear the primary impact of the defendant's act; rather, the plaintiff's relationship to the primary victim is diminished as a consequence.[179] Thus, loss of consortium may be regarded as a "secondary layer of tort liability superimposed upon the defendant's liability to the primary victim."[180] Yet it is not true in the law of torts that all those who predictably suffer loss as a consequence of the tortious act will be compensated.[181] Rather, it is for the judiciary to carefully demarcate the line of liability.

Generally, the degree of liability is dependent upon the degree of culpability. Thus, because strict tort liability is not premised upon culpability,[182] the scope of the manufacturer's liability should arguably be less in a strict products liability action than in a negligence action. Yet determining the appropriate diminution of the manufacturer's liability must involve a delicate balancing of interests. A judicial analysis and assessment of the relevant policies in light of society's current concerns about the overextension of the strict tort liability theory is required. The question must be whether the liability should encompass a relational and intangible injury such as loss of consortium.

The underlying purpose of the manufacturer's strict liability in tort should be a primary consideration. The legal theory was a judicial response to the increasing number of injuries caused by defective products and the corresponding difficulty of proving the negligence of a remote manufacturer.[183] The relaxation of the fault requirement was justified by strong policy considerations.[184] However, it does not follow that those

---

177. Rodriguez v. Bethlehem Steel Corp., 12 Cal. 3d 382, 400, 525 P.2d 669, ___, 115 Cal. Rptr. 765, 776 (1974).

178. Sizemore v. Smock, 430 Mich. 283, 294, 422 N.W.2d 666, 671 (1988).

179. Berger v. Weber, 411 Mich. 1, ____, 303 N.W.2d 424, 435 (1979) (Levin, J., dissenting).

180. *Id.*

181. *See Sizemore*, 430 Mich. at 292-93, 422 N.W.2d at 671 (the general rule in tort law is that a tortfeasor's liability only extends to an obligation to compensate the person directly injured); Norwest v. Presbyterian Intercommunity Hosp., 293 Or. 543, 652 P.2d 318, 333 (1982) (Tanzer, J., concurring) (court was hesitant to create an entitlement that is so unpredictable, formless and limitless). *Accord* Deloach v. Companhia de Navegacao Lloyd Brasileiro, 782 F.2d 438, 441 (3d Cir. 1986) (loss of consortium is somewhat an anomaly in the law of tort).

182. *See* RESTATEMENT, *supra* note 46, § 402A comments a and m; *see also* Shepard v. Superior Court, 76 Cal. App. 3d 16, 26, 142 Cal Rptr. 612, 618 (1977) (Kane, J., dissenting).

183. *See supra* notes 33-35 and accompanying text.

184. *See supra* notes 37-43 and accompanying text.

rationales adequately support the imposition of all the recoveries available under a negligence theory upon a manufacturer in a strict tort liability action. As a result, the validity of loss of consortium recovery under a negligence theory is not sufficient justification for the same recovery under a strict tort liability theory.

In addition, the burden of payment of awards for loss of consortium must realistically be borne by the public generally through increased insurance premiums or, otherwise, in the enhanced danger that accrues from the greater number of people or companies choosing to go without any insurance.[185] As one court aptly noted:

> There is a limit to the range of injuries and the dollar amount of recovery which can be spread across society through the interaction of the tort litigation and insurance systems. Increasing the load on the reparation system by recognizing causes of action in secondary tort victims in addition to the primary victim's action must increase insurance premiums, decrease participation in the system by marginal insureds, and perhaps decrease the amount that an insurer will willingly pay to the primary victim, thereby increasing litigation.[186]

Therefore, the inadequacy of monetary damages to make whole the loss suffered,[187] considered in light of the social costs of paying such awards, constitutes strong reason for denying the loss of consortium recovery in section 402A actions.

Moreover, it must be emphasized that section 402A was not intended to be an exclusive remedy.[188] The drafters of section 402A intended to impose a "special" liability upon the manufacturer or seller.[189] Other sections within Chapter 14 of the Restatement, in which section 402A appears, impose liability for negligence upon suppliers of chattels. For example, section 388 deals with chattels known to be dangerous for intended use;[190] section 395 imposes liability for the negligent manufacture of goods dangerous unless carefully made;[191] and section 398 addresses

---

185. Borer v. American Airlines, Inc., 19 Cal. 3d 441, 447, 563 P.2d 858, ____, 138 Cal. Rptr 302, 306 (1977).

186. Berger v. Weber, 411 Mich 1, ____, 303 N.W.2d 424, 438 (1981) (Levin, J., dissenting).

187. *Borer*, 19 Cal. 3d at 447, 563 P.2d at ____, 138 Cal. Rptr. at 306 (1977). *But cf.* Ueland v. Reynolds Metal Co., 103 Wash. 2d 131, 691 P.2d 190, 194 (1984).

188. *See* RESTATEMENT, *supra* note 46, § 402A comment a.

189. The title of § 402A is: Special Liability of Seller of Product for Physical Harm to User or Consumer. *See* RESTATEMENT, *supra* note 46, § 402A.

190. *See* RESTATEMENT (SECOND) OF TORTS § 388 (1977).

191. *See id.* 395.

the manufacture of goods made under a dangerous plan or design.[192] Each of these negligence sections recognizes the non-privity rule of *MacPherson*,[193] yet the basis of each is clearly fault.

Thus, precluding loss of consortium damages in strict products liability cases would not necessarily prevent redress of the legal wrong. It would merely require the loss of consortium plaintiff to be able to allege a different cause of action against the manufacturer defendant; and this would not generally affect recovery. An economic analysis of products liability law by Posner and Landes indicated that an efficient solution is just as likely to be reached under negligence as strict liability.[194] Further, even Prosser noted that there "is not one case in a hundred in which strict products liability would result in recovery where negligence does not."[195]

Thus, a judicial analysis should consider the relational nature of the loss of consortium cause of action when balancing against the social ramifications and the fact that alternative theories are available for recovery. Arguably, there are now sufficient countervailing policies to preclude the claim within a strict tort liability action.

IV. Conclusion

A perpetual challenge for the law is to limit redress for legal wrongs to a controllable degree.[196] This is especially true when the liability is the result of policy determinations and not premised upon the actor's culpable conduct. Yet the judiciary has not sufficiently met that challenge in the area of loss of consortium recovery in strict products liability actions against manufacturers and sellers.

The pronouncement in section 693 of the Restatement (Second) of Torts that loss of consortium is within the scope of the manufacturer's responsibility is not adequately supported by case law or detailed judicial analysis.[197] In addition, because the express provisions of section 402A

---

192. *See id.* 398.

193. *See* MacPherson v. Buick Motor Co., 217 N.Y. 382, 111 N.E. 1050 (1916).

194. Landes & Posner, *supra* note 88, at 541. Yet the economic argument for strict products liability to *physically injured* bystanders is stronger than the argument for strict liability to people in the chain of title. *Id.* at 551; *see also* Posner, *Strict Liability: A Comment*, 2 J. Legal Stud. 205, 207 (1973).

195. Prosser, *Assault Upon the Citadel*, *supra* note 3, at 1114.

196. An often quoted opinion noted:

> Every injury has ramifying consequences, like the ripplings of the waters, without end. The problem for the law is to limit the legal consequences of wrongs to a controllable degree.

Tobin v. Grossman, 24 N.Y.2d 609, 619, 249 N.E.2d 419, 424, 301 N.Y.S.2d 554, 561 (1969).

197. *See supra* notes 90-118 and accompanying text.

proscribe recovery to a plaintiff without physical injury,[198] it is reasonable to assert that recovery for a relational injury such as loss of consortium was not considered appropriate. Moreover, recovery for loss of consortium is not validated even if the plaintiff could be deemed analogous to a products liability bystander.[199] Yet the courts have consistently permitted the loss of consortium claim under a strict tort theory.

Again, in determining the parameters of the manufacturer's strict tort liability there must be a proper balance between the need for adequate recovery and the survival of a viable enterprise.[200] At the same time, there must be a recognition that the risk of harm from the loss of a loved one is pervasive. As one court noted, this is a risk of living and bearing children.[201] Thus, it is a risk that individuals must partially assume. Considering the fact that loss of consortium damages would continue to be compensable in products liability actions based on negligence or breach of warranty theories,[202] the preclusion of the cause of action under a strict tort liability theory would not be unreasonable. Accordingly, the time is appropriate for a judicial examination of loss of consortium recovery in a strict products liability action under section 402A of the Restatement (Second) of Torts.

Karen A. Jordan

---

198. *See supra* notes 119-33 and accompanying text.

199. *See supra* notes 134-60 and accompanying text.

200. *See, e.g.*, Helene Curtis Industries, Inc. v. Pruitt, 385 F.2d 841, 862 (5th Cir. 1967).

201. *See Tobin*, 24 N.Y.2d at 619, 249 N.E.2d at 424, 301 N.Y.S.2d at 562.

202. *See supra* notes 188-95 and accompanying text.

The Test for Patent Infringement Under the Doctrine of Equivalents After Pennwalt v. Durand-Wayland

The scope of a patented invention is defined by the claims in its application. When each element of a claim has been copied by an accused device, the claim is said to "read upon" the new invention, and is literally infringed by that invention.[1] However, this type of infringement is rare. Far more often it is the case that the new invention is substantially similar to the claim and thereby infringes it under the doctrine of equivalents.

Recently there has been much controversy over the judicial standard for infringement analyses under the doctrine of equivalents. Arguably, there are three distinct standards. The first type is known as the "element-by-element" analysis, and consists of comparing each element of the patented claim with the accused device to determine if the accused device infringes upon the patent.[2] The second type of analysis is a version of the "invention as a whole" approach whereby the accused device is compared with the claimed invention as a whole to determine if it infringes upon that invention.[3] Proponents of either standard have heatedly argued the validity of their positions without recognizing one simple concept: the element-by-element standard and the invention as a whole approach are not mutually exclusive. The last type of infringement analysis is an integration of both the element-by-element and invention as a whole standards. The conflict between the different analyses of infringement under the doctrine of equivalents finally culminated in an *en banc* decision of the Court of Appeals of the Federal Circuit, *Pennwalt Corp. v. Durand-Wayland, Inc.*[4]

This Note focuses on the confusion in pre-*Pennwalt* cases that resulted in different analyses of infringement under the doctrine of equivalents, the *Pennwalt* decision, and the framework for infringement analyses after *Pennwalt*. A close examination of case precedent discloses that the correct framework is an integration of both the element-by-element and the invention as a whole standards. This Note concludes with an analysis protocol which should be followed by all federal courts when determining infringement under the doctrine of equivalents.

---

1. I. Kayton, Patent Practice, §§ 2-12, 13 (1985).
2. *See infra* text accompanying notes 58-62, 87-89.
3. *See infra* text accompanying notes 115-16.
4. 833 F.2d 931 (Fed. Cir. 1987), *cert. denied*, 108 S.Ct. 1226 (1988), *cert. denied*, 108 S.Ct. 1474 (1988).

I. Basic Principles of Patent Infringement

A. *Literal Infringement and Infringement Under the Doctrine of Equivalents*

In a landmark decision, *Graver Tank & Manufacturing Co. v. Linde Air Products*,[5] the United States Supreme Court enunciated the prima facie case for infringement under the doctrine of equivalents—the two inventions must "do the same work in substantially the same way, and accomplish substantially the same result."[6] Through the years, variations in terminology have resulted in a variety of combinations of phrases which express the test for equivalency: "(a) substantially the same result, or end function, or does substantially the same work; (b) in substantially the same way, or manner, or mode of operation, and; (c) using substantially the same means or structure."[7] Whichever combination of the above phrases is used, the meaning is the same as originally stated in *Graver Tank*—equivalence must be based upon a substantial identity of result, function, and means.[8]

---

5. 339 U.S. 605 (1950). The doctrine of equivalents has earlier roots, dating as far back as the 1800s. *See* Winans v. Denmead, 56 U.S. 330 (1853). *See also* Morley Machine Co. v. Lancaster, 129 U.S. 263, 273 (1888) ("Where an invention is one of a primary character, and the mechanical functions performed by the machine are, as a whole, entirely new, all subsequent machines which employ substantially the same means to accomplish the same result are infringements, although the subsequent machine may contain improvements in the separate mechanisms which go to make up the machine.") However, *Graver Tank* is the Supreme Court decision which is the most often cited in current cases for the enunciation of the doctrine of equivalents.

6. *Id.* at 608. *See also* Swanson, *A Discussion of the Application of the Doctrine of Equivalents in the Graver v. Linde Case*, 33 J. Pat. Off. Soc'y 19 (1951); Tilton, *The Doctrine of Equivalents in Patent Cases*, 32 J. Pat. Off. Soc'y 861 (1950).

7. Arnold & Lynch, *Infringement Of Inventions*, The John Marshall Law School 28th Annual Conference: More Developments in Intellectual Property Law 14 (1984).

8. *Id.* at 14-16. *But see* Harris, *Three Ambiguities Of The Doctrine Of Equivalents In The Federal Circuit*, 69 J. Pat. Off. Soc'y 91 (1987). One of the ambiguities which the Harris article addresses is the different terminology the Federal Circuit has used in several opinions, stating the test for equivalence as either "substantially the same result" or "the same result," or even both in the same opinion. The Harris article finds that "the difference between the two formulations of the basic test is important in those situations in which the accused device achieves substantially the same result, but does *not* achieve the *same* result as the patented item, and in which the accused device satisfies the 'function' and 'way' legs of the basic test." *Id.* at 93 (emphasis in original). This ambiguity is outside of the scope of this article, and perhaps can be best summed up by stating, "Only legal theorists worry over the semantic differences, because what the courts are trying to do is articulate their words for 'substantially the same invention.'" Arnold & Lynch, *supra* note 7, at 16.

A discussion of both literal infringement and infringement under the doctrine of equivalents using hypothetical claims will be helpful for a more complete understanding of the fundamental principles of infringement involved. There are generally two types of clauses in a claim: those using structural language and those using functional (also known as "means-plus function") language.[9] The following hypothetical claims for a table illustrate both types of clauses.

(1) A device, comprising:

> a flat surface; at least three legs perpendicular to the flat surface; and screws securing each leg to the flat surface.

(2) A device, comprising:

> a flat surface; at least three legs perpendicular to the flat surface; and means for securing each leg to the flat surface.

The first claim uses structural language, while the second claim uses functional language.

The important difference between the two claims when determining whether the claims have been literally infringed is the leeway statutorily granted to functional clauses by the Patent Act of 1952.[10] Section 112 of the Act provides in pertinent part that:

> An element in a claim for a combination may be expressed as a means or step for performing a specified function without the recital of structure, material, or acts in support thereof, and such claim shall be construed to cover the corresponding structure, material, or acts described in the specification and *equivalents thereof.*[11]

This statute not only permits an element to be set forth in functional language, but it also provides that the claim will be interpreted to include equivalents of that functional element.[12] Therefore, in determining whether an accused device literally infringes a claim, functional language is construed to include equivalents of that element. Structural language is not permitted this leeway in literal infringement.[13]

---

9. KAYTON, *supra* note 1, at §§ 2-22, 23.

10. Ch. 950, 66 Stat. 797 (codified at 35 U.S.C. §§ 100-135 (1982)).

11. 35 U.S.C. 112 (6) (1982) (emphasis added).

12. KAYTON, *supra* note 1, at §§ 2-23.

13. Atlas Powder Co. v. E. I. Du Pont De Nemours & Co., 750 F.2d 1569 (Fed. Cir. 1984). (Substitution of a certain element in the claim ("water-in-oil type emulsifying agent") for a similar element in the accused product ("oil-in-water emulsifying agent") avoided literal infringement. Both elements are in structural, rather than functional, language. The substitution, however, did not preclude a finding of infringement under the doctrine of equivalents).

To illustrate this point, consider the two hypothetical claims. An accused device which has legs attached to a flat surface by weld, glue, rivets, or joints, does not literally infringe the claim containing structural language. However, it does literally infringe the claim containing functional language, because the "means-plus function" clause is deemed to include equivalent methods of attaching the legs to the flat surface.[14]

The doctrine of equivalents is a creature of equity which expands upon this premise and allows a finding of infringement where the accused product does not literally infringe upon the claim, but is substantially equivalent to the entire claim.[15] For example, incorporating the *Graver Tank* test to the above hypotheticals, if a court finds that an accused device which differs from the structural claim only by using rivets rather than screws "[does] the same work in substantially the same way, and accomplish[es] substantially the same result,"[16] then that device infringes under the doctrine of equivalents. Although the doctrine of equivalents is related to statutory "means-plus function" equivalence, statutory equivalence is determinative only on the issue of literal infringement, not infringement under the doctrine of equivalents.[17] "Section 112, paragraph 6, plays no role in determining whether an equivalent function is performed by the accused device under the doctrine of equivalents."[18]

The rationale for the doctrine of equivalents is to prevent infringers from practicing fraud on a patent.[19] As the Court explained in *Graver Tank*:

> [T]o permit imitation of a patented invention which does not copy every literal detail would be to convert the protection of

---

14. Kayton, *supra* note 1, at §§ 2-23, 24.
15. *Id.* at 2-25.
16. 339 U.S. at 608.
17. Pennwalt Corp. v. Durand-Wayland, Inc., 833 F.2d 931, 933-4 (Fed. Cir. 1987).
18. *Id.* at 934. *See also* D.M.I., Inc. v. Deere & Co., 755 F.2d 1570, 1575 (Fed. Cir. 1985).

> [T]he word 'equivalent' in 112 should not be confused, as it apparently was here, with the 'doctrine of equivalents.' In applying the doctrine of equivalents, the fact finder must determine the range of equivalents to which the claimed invention is entitled, in light of the prosecution history, the pioneer-non-pioneer status of the invention, and the prior art. It must then be determined whether the entirety of the accused device or process is so 'substantially the same thing, used in substantially the same way, to achieve substantially the same result' as to fall within that range. In applying the 'means plus function' paragraph of 112, however, the sole question is whether the single means in the accused device which performs the function stated in the claim is the same as or an equivalent of the corresponding structure described in the patentee's specification as performing that function.

Id. at 1575 (citations omitted), *later appeal*, 802 F.2d 421 (Fed. Cir. 1986).

19. *Graver Tank*, 339 U.S. at 608.

> the patent grant into a hollow and useless thing. Such a limitation would leave room for —indeed encourage— the unscrupulous copyist to make unimportant and insubstantial changes and substitutions in the patent which, though adding nothing, would be enough to take the copied matter outside the claim, and hence outside the reach of law. One who seeks to pirate an invention, like one who seeks to pirate a copyrighted book or play may be expected to introduce minor variations to conceal and shelter the piracy. Outright and forthright duplication is a dull and very rare type of infringement. To prohibit no other would place the inventor at the mercy of verbalism and would be subordinating substance to form. It would deprive him of the benefit of his invention and would foster concealment rather than disclosure of inventions, which is one of the primary purposes of the patent system.[20]

The doctrine of equivalents is a device used to expand the meaning of the claims in a patent in order to promote the policy of encouraging progress in the arts by protecting the patent owner.[21] The limitations which courts place on the doctrine of equivalents derive from the policy of giving competitors fair notice of what the patent owner believes to be the boundaries of his invention, so that competitors may design around the patent without open vulnerability to infringement actions.[22]

---

20. *Id.* at 607. *See also* Royal Typewriter Co. v. Remington Rand, Inc., 168 F.2d 691, 692 (2d Cir. 1948) (Judge Hand stated that courts will apply the doctrine of equivalents to "temper unsparing logic and prevent an infringer from stealing the benefit of the invention." He described the doctrine of equivalents as an anomaly, but a necessary one, accepted by the courts to prevent inequity and injustice.), *cert. denied*, 335 U.S. 825 (1948), *reh'g denied*, 335 U.S. 864 (1948) and *petition denied* 82 U.S.P.Q. (BNA) 334 (2d Cir. 1949).

21. *Pennwalt*, 833 F.2d at 945 (Bennett, J., dissenting).

22. *Id. But cf.* Adelman and Francione, *The Doctrine of Equivalents in Patent Law: Questions that Pennwalt Did Not Answer*, 137 U. Pa. L. Rev. 673 (1989).

> The doctrine of equivalents is the primary (although not the exclusive) cause of the current uncertainty surrounding the scope of patent claims. This uncertainty has serious consequences. First, uncertainty about the scope of patent protection hinders both patent holders and potential defendants from assessing the possible outcome of litigation or from making other business decisions, such as the direction that research and development efforts should take. Second, a primary purpose of the protection of intellectual proprty is to encourage the production of inventions, literary works, and the like. Patent law in particular provides a claiming system to put other potential inventors on notice of the precise boundaries of the invention so that they may "design around" the patent other inventive efforts. The uncertainty generated by the doctrine of equivalents frustrates and chills the activities of these other inventors, who must be concerned about whether their efforts will be met by an infringement suit based on the

In light of the rationale underlying the doctrine, it is logical that the range of equivalents allowed under the doctrine of equivalents "depends upon and varies with the degree of invention."[23] The purpose of the doctrine of equivalents is to reward the inventor by protecting his rights to his patent.[24] This reward to the inventor in the form of protection of his patent through the application of the doctrine of equivalents is commensurate with the value of the inventor's contribution to the art.[25] Generally, there are three categories of ranges of equivalents: (a) pioneer inventions are entitled to a broad range of equivalents; (b) marked improvements are entitled to a substantial range of equivalents; and, (c) narrow improvements are entitled, at most, to a limited range of equivalents.[26] Therefore, when determining whether an accused device is substantially equivalent to the claim, and thus infringes it under the doctrine of equivalents, the court will allow the claim of a pioneer invention a much broader range of equivalents than the claim of an invention which is only a small advancement in a crowded field.

B. Procedure for Analysis of Patent Infringement

Courts traditionally follow a two-step approach to analyze patent infringement issues.[27] First, the claims are construed by the courts to determine their scope.[28] Second, the scope of the claims is then compared

---

> amorphous doctrine of equivalents. Third, the doctrine permits abusive infringement actions claiming that the defendant infringes under the doctrine of equivalents and that a jury must decide the correctness of the claim. The imperative to settle under these circumstances is almost overpowering. Fourth, due process concerns are potentially raised to the extent that pervasive and systemic uncertainty generated by the doctrine of equivalents destroys the ability of patent claims to provide fair notice, so that they effectively provide no notice.

Id. at 682 (citations omitted).

23. Continental Paper Bag Co. v. Eastern Paper Bag Co., 210 U.S. 405 (1908).

24. Nelson v. Batson, 322 F.2d 132, 135 (9th Cir. 1963).

25. *Id.* The *Nelson* case states:

> Since the purpose of the doctrine of equivalents is to give the inventor an opportunity to secure a just reward for his invention—an opportunity which he would otherwise be denied because of the failure of the language of his claim to include devices which were in fact the same as his own in function, means, and result —the degree of protection afforded beyond the language of the claims will vary directly with the value of the inventor's contribution to the art.

Id. at 135.

26. D. CHISUM, PATENTS, 18.04[2] at 18-36.1-42 (1986) (citations omitted). *See also* HAYES, THE NATURE OF PATENTABLE INVENTION, 174-181 (1948).

27. ZMI Corp. v. Cardiac Resuscitator Corp., 844 F.2d 1576, 1578 (Fed. Cir. 1988) (quoting Mannesmann Demag Corp. v. Engineered Metal Products Co., 793 F.2d 1279, 1282 (Fed. Cir. 1986)).

28. *Id.*

with the accused device to determine whether the scope of the claims encompasses the accused device.[29] Questions of literal infringement or infringement under the doctrine of equivalents arise under the second of the two steps, the comparison of the scope of the claims with the accused device. However, the first step, the determination of the scope of the claims, is integral to a proper analysis of infringement.[30] Therefore, in order to gain a more complete understanding of the infringement analysis process, a discussion of how courts determine the scope of a claim is essential.

Courts construe claims by inspecting the language of the claims themselves, the specification, the drawings, and the file wrapper.[31] The specification is helpful in determining the scope of the claims because "the words must be used in the same way in both the claims and the specification."[32] Patents may also contain drawings which can be useful as visual representations to "flesh out" the words used in the claims and specification.[33] The final aid which courts use in determining the scope of a patent is the file wrapper.

The file wrapper is a record of the prosecution history of the patent. It contains "the entire record of the proceedings in the Patent Office from the first application papers to the issued patent."[34] In claim construction, file wrapper estoppel (also known as prosecution history estoppel) is used to narrow the scope of the claims to the extent the claims had been narrowed during the prosecution history to avoid prior art rejection.[35] During the application process of a patent, the Patent Office may reject the application on the grounds that the invention is too broad, and encompasses prior art.[36] In order to have his patent accepted, the inventor will then amend, add, or cancel claims of his patent.[37] After making such changes, the inventor is estopped from claiming that the scope of the claims in his patent includes the subject matter that was surrendered during the prosecution history of his patent.[38]

---

29. *Id.*

30. *Id.* ("Improper claim construction, i. e., an improper determination of the scope of the claims, can distort the entire infringement analysis." Quoting Moeller v. Ionetics, Inc., 794 F.2d 653, 656 (Fed. Cir. 1986)).

31. Autogiro Co. v. United States, 384 F.2d 391 (Ct. Cl. 1967). *Cf.* SRI International v. Matsushita Electric Corp., 775 F.2d 1107, 1118 (Fed. Cir. 1985) ("A claim is construed in the light of the claim language, the other claims, the prior art, the prosecution history, and the specification, *not* in light of the accused device.") (emphasis in original).

32. *Autogiro Co.*, 384 F.2d at 397.

33. *Id.*

34. *Id.*

35. Kayton, *supra* note 1, at 2-27.

36. Chisum, *supra* note 26, at 18.05[2].

37. *Id.*

38. Hughes Aircraft Co. v. United States, 717 F.2d 1351 (Fed. Cir. 1983).

During an infringement action, file wrapper estoppel constitutes a powerful limitation on the doctrine of equivalents because the equivalents of the patent's claims may never include material that the patent owner surrendered during the prosecution history of his patent.[39]

After determining the scope of the claims, the courts will then proceed to the second step of an infringement analysis: comparison of the scope of the claims to the accused device to determine whether such device infringes upon the claims. Within this second step, the first issue is literal infringement.[40] If the accused device literally infringes the scope of the claims of the patent, then "infringement is made out and that is the end of it."[41] Only if the court finds that the accused device does not literally infringe upon the scope of the claim, does it then proceed to the issue of infringement under the doctrine of equivalents.[42]

How do courts apply the doctrine of equivalents when determining whether the accused device infringes upon the patented claims? Two schools of thought have developed with respect to the standard of applicability for the doctrine of equivalents. In cases such as *Hughes Aircraft Co. v. United States*[43] and *Texas Instruments Inc. v. International Trade Commission*,[44] the Court of Appeals for the Federal Circuit (hereinafter "Federal Circuit") held that the doctrine of equivalents was to be applied to the "invention as a whole."[45] These cases appear to contradict other cases such as *Lemelson v. United States*[46] and *ACS Hospital Systems, Inc. v. Montefiore Hospital*,[47] in which the Federal Circuit applied the doctrine of equivalents on an element-by-element basis and found that the accused product did not infringe upon the

---

39. Townsend Engineering v. HiTec, 829 F.2d 1086 (Fed. Cir. 1987), *later proceeding*, 117 F.R.D. 612 (N.D. Ill. 1987).

40. Graver Tank & Mfg. Co. v. Linde Air Products, 339 U.S. 605, 607 (1950).

41. *Id. But see* Phillips Petroleum v. U. S. Steel Corp., 6 U.S.P.Q.2d (BNA) 1065, 1123 (D. Del. 1987). (Even if the accused product falls within the literal language of the claim, the reverse doctrine of equivalents may preclude a finding of infringement. Under the reverse doctrine of equivalents, "where the device is so far changed in principle from a patented article that it performs the same or similar function in a substantially different way but nevertheless falls within the literal words of the claim," the doctrine is used to restrict the claim, thus preventing a finding of literal infringement. Quoting SRI International v. Matsushita Electric Corp., 775 F.2d 1107, 1123 (Fed. Cir. 1985)), *aff'd*, 865 F.2d 1247 (Fed. Cir. 1989).

42. *Id.*

43. 717 F.2d 1351 (Fed. Cir. 1983).

44. 805 F.2d 1558 (Fed. Cir. 1986), *reh'g denied*, 846 F.2d 1369 (Fed. Cir. 1988), *reh'g denied*, (en banc) 7 U.S.P.Q.2d 1414 (Fed. Cir. 1988).

45. *See supra* notes 43 and 44.

46. 752 F.2d 1538 (Fed. Cir. 1985).

47. 732 F.2d 1572 (Fed. Cir. 1984).

patented device because the accused product lacked an element present in the patented device.[48]

Much of the confusion that has been generated over the standard of applicability for the doctrine of equivalents stems from the assumption by many practioners that these two standards are mutually exclusive—that to choose one is to reject the other. This assumption is not correct. Both standards should be integrated into a single framework in which to apply the doctrine of equivalents. In *Perkin-Elmer Corp. v. Westinghouse Electric Corp.*,[49] the Federal Circuit hinted at this framework when it applied an element-by-element analysis while simultaneously stating that it is "legal error not to 'apply the doctrine of equivalents to the claimed invention as a whole.'"[50]

The conflict between the proponents of a single framework, incorporating both the element-by-element analysis and the invention as a whole analysis, and those who chose one standard to the exclusion of the other finally culminated in *Pennwalt Corp. v. Durand-Wayland, Inc.*[51] The majority of the Federal Circuit in *Pennwalt* held in an *en banc* decision, that the standard of application for the doctrine of equivalents incorporates both the element-by-element standard and the invention as a whole standard.[52] The remainder of this Note discusses the standard of applicability for the doctrine of equivalents as it evolved before *Pennwalt*, the *Pennwalt* decision, and what remains of the doctrine of equivalents after *Pennwalt*. It will be shown that courts have implicitly, if not explicitly, adopted both standards (the invention as a whole standard and the element-by-element standard) into a single framework in which to analyze infringement issues. The evolution of these two distinct standards results in a single workable framework that fulfills the goals of patent law by protecting the patent owner's right in his

---

48. *See supra* notes 46 and 47.

49. 822 F.2d 1528 (Fed. Cir. 1987).

50. Id. at 1532 (quoting Hughes Aircraft Co. v. United States, 717 F.2d 1351, 1364 (Fed. Cir. 1983)). *See also* Nieman, *The Federal Circuit Resolves Ambiguities in the Doctrine of Equivalents*, 70 J. PAT. OFF. SOC'Y 91, 155 (1987) ("Legitimate distress over the Federal Circuit's 'new infringement analysis' was short-lived. The author of *Hughes Aircraft*, Chief Judge Markey, clarified that opinion's use of the 'invention as a whole' concept in *Perkin-Elmer Corp. v. Westinghouse Electric Corp.*"). *But see* Hantman, *Doctrine of Equivalents*, 70 J. PAT. OFF. SOC'Y 511, 548 (1988) ("Apparently, all these 'doctrines of equivalents' were too much for the Federal Circuit. In *Perkin-Elmer Corp. v. Westinghouse Electric Corp.*, the Federal Circuit appeared to back away from the doctrine of equivalents to the claimed invention as a whole. The Federal Circuit said that every element of a claim is material and every element or its substantial equivalent must be present in the infringing structure.")

51. 833 F.2d 931 (Fed. Cir. 1987), *cert. denied*, 108 S. Ct. 1226 (1988), *cert. denied*, 108 S. Ct. 1474 (1988).

52. 833 F.2d 931, 935.

invention while giving his competitors notice of what constitutes his patent so that they may compete freely with him, without fear of infringing upon his patent.

II. The Evolution of a Single Framework from the "Invention as a Whole" Standard and the "Element-by-Element" Standard Prior to *Pennwalt*

Until recently, case law concerning the doctrine of equivalents has not expressly addressed whether the standard of applicability is as an invention as a whole, or element-by-element, or a single framework incorporating both standards. Two extensive essays covering the development of the standard of applicability have reviewed case precedent since the Supreme Court's first discussion of the doctrine of equivalents in *Winans v. Denmead*[53] in 1853 to the *Graver Tank*[54] decision in 1950, and have arrived at entirely opposite conclusions. An article written by Hantman takes the viewpoint that not only does the invention as a whole standard have no legal basis in case precedent, but it goes one step further and concludes that the expansive doctrine of equivalents expounded in *Graver Tank* was "an anomaly that should have been overruled with the passage of 35 U.S.C 112(6)" of the Patent Act of 1952.[55] Judge Newman's dissenting opinion in *Pennwalt* discusses the

---

53. 56 U.S. 330 (1853).

54. 339 U.S. 605 (1950).

55. Hantman, *supra* note 50, at 551.

> In addition to having no legal basis, the 'invention as a whole' approach to the doctrine of equivalents is not a workable doctrine. The approach provides no basis for analysis for infringement determination. It is ambiguous, uncertain and guaranteed to provide much litigation for the courts. Quite simply, it is not understandable. The *Pennwalt* expansive doctrine of equivalents applied on an element by element basis is a retreat from Hughes and *Texas Instruments*. However, the *Pennwalt* decision did not go far enough. It should also have retreated from the doctrine of equivalents of *Graver Tank*. The *Graver Tank* decision contradicted the previous one hundred years of Supreme Court precedents and the *Pennwalt* court should have considered it overruled by the last paragraph of Section 112, 35 U.S.C. 112 (6), of the Patent Act of 1952.

Cf. Jessup, *The Doctrine of Equivalents*, 54 J. Pat. Off. Soc'y 248, 270 (1972) ("The doctrine [of equivalents] is basically unsound and unfair to the patentee's competitors. Its continued existence is a reproach of the patent bar."); Klitzke, *Patent Law: Equivalency and Validity in the Seventh Circuit*, Intellectual Prop. L. Rev. 49, 53 (1979) ("The doctrine of equivalents is a superfluous gloss on the patent law. The breadth of the generic terms allowed in the claims is the measure of the degree of advancement in the art. Reviewing courts must be cognizant of two facts: (1) the patentee will always generously expand the scope of the discovery in the specification, and (2) regardless of this, the test of patent scope ultimately must be the terms allowed by the Patent Office in the issued claims. To excessively extend the scope of the claims by the doctrine of equivalents may neutralize many months of negotiation in the patent prosecution process.")

same case precedent as the Hantman article. However, the similarity ends there. Newman's opinion is a strong proponent of the invention as a whole standard, and rejects the element-by-element analysis as "contrary to the overwhelming body of precedent."[56] The inapposite conclusions which these two well-reasoned essays have reached demonstrate how easy it was for two separate standards of applicability to develop and that a review of case precedent prior to recent years is not determinative of this issue.[57] Rather, a review of more recent cases which detail the evolution of both distinct standards into a unified framework is far more helpful to a determination of the state of case precedent on this issue prior to the *Pennwalt* decision.

A. *Interdent: An Element-by-Element Approach*

In a 1976 decision, *Interdent Corp. v. United States*,[58] the Court of Claims (whose decisions are binding on the Federal Circuit) applied an element-by-element analysis, and held that the accused product, a Water Pik surgical irrigator, did not infringe on the claim of another surgical irrigator because the accused product lacked the composite switch and valve set forth in the claim of the original device.[59] Instead, the accused product only used a single switch to turn the unit on, and did not use a valve to control the fluid flow through the outlet conduit.[60] The lack of an element in the claim of the original device, the valve, prevented a finding of infringement under the doctrine of equivalents. The court in *Interdent* held that "each element of a patented combination is considered to be material and essential. Thus, the omission of any one of the elements of the claimed combination avoids infringement."[61] This is known as the "all elements" rule and typifies the element-by-element approach.[62]

However, even if the court in *Interdent* had applied the invention as a whole standard of applicability of the doctrine of equivalents, it is doubtful that the court would have arrived at a different conclusion. During the prosecution of the original patent, the claim was accepted by the Patent Office only after it had distinguished itself from existing patents by showing that it had improved upon existing devices by com-

---

56. 833 F.2d at 974 (Newman, J., dissenting commentary).

57. Hantman, *supra* note 50, at 551. ("However, the dissenting opinion in the *Pennwalt* case shows that one may take a different view of the precedents and arrive at a different conclusion.")

58. 187 U.S.P.Q. (BNA) 523 (Ct. Cl. 1975), *aff'd per curiam*, 531 F.2d 547 (1976).

59. *Id.*

60. 531 F.2d at 553.

61. *Id.* at 552.

62. *Id.*

bining a novel valve and switch structure. The court in *Interdent* stated "in instances such as those at bar, where a patent depends for its novelty on a single feature, courts have been reluctant to expand the coverage of a claim by the doctrine of equivalents to cover a device which lacks that single feature."[63] This limitation on the range of equivalents where the patentee has narrowed his claims in order to avoid the prior art is an example of file wrapper estoppel.[64] It constitutes a valid restriction on the range of equivalents that may be considered when applying the doctrine of equivalents. Thus, even if the court in *Interdent* had viewed the invention as a whole, the original patentee would still have been estopped from claiming infringement where the accused product lacked the one feature that constituted the point of novelty of the original patent.

B. Hughes Aircraft: The Federal Circuit Expresses the Invention as a Whole Standard

A 1983 decision written by Chief Judge Markey, *Hughes Aircraft Co. v. United States*,[65] was the first time the Federal Circuit expressly set forth the invention as a whole standard.[66] The court held that it was reversible error to fail to apply the doctrine of equivalents to the invention as a whole.[67] In *Hughes Aircraft*, the Federal Circuit addressed a claim of infringement concerning a patented satellite which had a unique attitude control system.[68] The claims included a means for transmitting an orientation signal to a fixed external coordinate system (earth) and a means for receiving control signals from that external location.[69] The satellite then used the control signals from earth to reorient itself by firing corrective jets.[70] The accused satellite was equivalent to the original one except that it did not transmit an orientation signal to earth, but rather supplied the orientation signal to an on-board computer which then used a control signal from earth to reorient the satellite.[71]

On the issue of infringement under the doctrine of equivalents, the trial court held that because there was "no obvious or exact equivalent" of two elements of the original claim, including the means for transmitting

---

63. *Id.*

64. *Id.* at 525, 551. *See supra* text accompanying notes 34-39 for a discussion of file wrapper estoppel.

65. 717 F.2d 1351 (Fed. Cir. 1983).

66. *Id.*

67. *Id.* at 1364.

68. *Id.* at 1353.

69. *Id.* at 1355.

70. *Id.* at 1354.

71. *Id.* at 1363-64.

an orientation signal to earth, the accused satellite did not infringe the claims.[72] The Federal Circuit reversed the holding of the trial court, and found that the two satellites were substantially equivalent.[73] In enunciating its decision, the Federal Circuit made the following statement, setting forth the invention as a whole requirement to an infringement analysis under the doctrine of equivalents. "The failure to apply the doctrine of equivalents to the claimed invention as a whole, and the accompanying demand for 'obvious and exact' equivalents of two elements the presence of which would have effectively produced literal infringement, was error."[74] With this single, seemingly innocent sentence, the Federal Circuit opened up a Pandora's box of future misconstructions and ambiguities. All recent cases which assert that the invention as a whole analysis is the only correct approach to determine infringement under the doctrine of equivalents, do so citing the above sentence from *Hughes Aircraft*.[75]

The above statement from the Federal Circuit in *Hughes Aircraft* actually refers to two distinct errors on the part of the trial court.[76] The first was the trial court's error in requiring "obvious and exact" equivalents for the two means-plus function clauses that were present in the original patent but were missing from the accused satellite.[77] The Federal Circuit correctly identified the trial court's requirement of "obvious and exact" equivalents as a more stringent standard than the "substantially the same function, way, and result" test for the doctrine of equivalents under *Graver Tank*.[78] The "obvious and exact" equivalents test imposes a more onerous burden on the patent owner to show infringement under the doctrine of equivalents than was previously mandated by case precedent. The Federal Circuit held that the trial court's use of the "obvious and exact" equivalents test was no more than a redundant literal infringement inquiry, and not dispositive of the issue of infringement under the doctrine of equivalents.[79] Moreover, the Federal Circuit was puzzled by why the "obvious and exact" equivalents test was not satisfied by evidence of the accused satellite's use of an on-

---

72. *Id.* at 1364.
73. *Id.* at 1366.
74. *Id.*
75. *See, e.g.*, Pennwalt Corp. v. Durand-Wayland, Inc., 833 F.2d 931, 940-42 (Fed. Cir. 1987) (Bennett, J., dissenting), 833 F.2d at 972 (Newman, J., dissenting commentary); Texas Instruments Inc. v. Int'l Trade Comm'n, 805 F.2d 1558, 1569-70 (Fed. Cir. 1986).
76. 717 F.2d at 1364.
77. *Id.*
78. *Id.*
79. *Id.* ("However the phrase 'obvious and exact equivalents may be defined, it was effectively and improperly applied here as a substitute for literal infringement, the absence of which was conceded." *Id.* at 1364.)

board computer to perform all of the functions of the two disputed means-plus function claims in the original patent.[80]

The second of the two distinct trial court errors was its failure to apply the doctrine of equivalents to the invention as a whole.[81] Much of the controversy regarding the standard of applicability for the doctrine of equivalents stems from the Federal Circuit's expression in *Hughes Aircraft* of the invention as a whole standard. As the Federal Circuit later clarified in *Perkin-Elmer Corp. v. Westinghouse Electric Corp.*,[82] the invention as a whole standard merely requires that each element be viewed in the context of the entire patent. There is no language in the *Hughes Aircraft* opinion that indicates that the invention as a whole analysis excludes the element-by-element approach.[83]

In fact, after stating that the trial court erred in failing to apply the doctrine of equivalents to the invention as a whole, the Federal Circuit then proceeded to compare each of the elements in the original patent to the accused satellite to determine if the elements or their equivalents were present in the accused device.[84] This is the heart of an element-by-element analysis. In comparing each of the disputed elements of the original patent to the accused satellite, the Federal Circuit viewed each element in the context of the entire claim to determine its appropriate range of equivalents.[85] This is what the court in *Hughes Aircraft* meant by an invention as a whole approach. The Federal Circuit's use of an element-by-element analysis in conjunction with the invention as a whole standard shows that the "invention as a whole" statement[86] was not meant to embody the sole type of analysis for infringement under the doctrine of equivalents. Rather, it was meant to express further limitations which courts had already implicitly applied on the element-by-element analysis.

C. ACS Hospital and Lemelson: The "All Elements" Rule After Hughes Aircraft

The "all elements" rule states that where "the accused device omits an element of the claimed invention, the accused device may therefore be viewed as failing to function in 'substantially the same way' as the

---

80. *Id.*

81. *Id.*

82. 822 F.2d 1528 (Fed. Cir. 1987). *See infra* text accompanying notes 123-28.

83. 717 F.2d 1351 (Fed. Cir. 1983).

84. *Id.* at 1364-66.

85. *Id. See supra* text accompanying notes 23-26, for a discussion of the range of equivalents to be afforded a claim under the doctrine of equivalents.

86. *Id.* at 1364.

claimed invention.''[87] This rule is based upon two interrelated premises. First, each element of a claim is material and essential.[88] Second, because each element is essential to the claim, the omission of any element in the claim prevents a finding of infringement under the doctrine of equivalents.[89] The ''all elements'' rule comprises the backbone of any element-by-element analysis, for if the accused device does not contain one of the elements of the original patent (or its equivalents), then it does not infringe upon that patent.

The application of the ''all elements'' rule to an infringement analysis under the doctrine of equivalents does not preclude the application of the invention as a whole standard. As shown by *Hughes Aircraft*, the invention as a whole standard acts as a limitation on how the element-by-element analysis is applied.[90] Two cases decided by the Federal Circuit after *Hughes Aircraft* affirm the use of a single framework incorporating both standards.

In 1984, the Federal Circuit, in an opinion written by Judge Smith, held in *ACS Hospital Systems, Inc. v. Montefiore Hospital*[91] that the accused device did not infringe upon the patented rental television system because it did not have one of the elements of the claim—a function for overriding a locked key switch.[92] In reaching a finding of non-infringement based on the absence in the accused device of an element in the original patent, the Federal Circuit's analysis typified the element-by-element approach. However, the court also implicitly applied the invention as a whole approach in its determination of the range of equivalents of the missing element.

The original patent only contained three elements: a key operated actuating switch, an override switch, and a signal light to indicate that the override switch has been actuated.[93] The Federal Circuit found that the accused device did not contain an override switch.[94] In construing the range of equivalents, the Federal Circuit viewed the override switch in the context of the entire claim of the original patent. This is the classic invention as a whole approach. Here, where the patented television rental system constitutes only a minor advancement over prior art,[95] it

---

87. Harris, *supra* note 8, at 98. *See also* Pennwalt Corp. v. Durand-Wayland, Inc., 833 F.2d 831, 949-52 (Fed. Cir. 1987) (Nies, J., additional views).

88. Interdent Corp. v. United States, 531 F.2d at 552 (Ct. Cl. 1976).

89. *Id.*

90. 717 F.2d 1351 (Fed. Cir. 1983).

91. 732 F.2d 1572 (Fed. Cir. 1984).

92. *Id.*

93. *Id.* at 1574.

94. *Id.* at 1579-81.

95. *Id.* (The trial court held that the original patent was invalid for obviousness

is accorded a narrow range of equivalents. Also, it may be reasoned as a general proposition that when the disputed element is one out of a few total elements in the claim, this element is more important to the patented combination of elements than when it comprises only one element out of a large number of elements in the claim. Logically, the more important the element is to the patented combination, the narrower should be the range of equivalents accorded to that element.[96] As a result of the narrow range of equivalents for the override switch, the court did not find infringement of the original patent under the doctrine of equivalents.

In 1985, the Federal Circuit again applied an integrated analysis of both the element-by-element and invention as a whole standards, this time in the case of *Lemelson v. United States*.[97] In *Lemelson*, the Federal Circuit, through an opinion written by Judge Baldwin, affirmed the trial court's finding of non-infringement where an element (manipulation means) of the original claim (surface sensing apparatus) was missing in the accused device.[98] Citing *Interdent Corp. v. United States*,[99] the court in *Lemelson* stated that "[i]t is also well settled that each element of a claim is material and essential, and that in order for a court to find infringement, the plaintiff must show the presence of every element or its substantial equivalent in the accused device."[100] This pronouncement of the "all elements" rule in *Lemelson* reaffirms the Federal Circuit's use of the element-by-element analysis in infringement issues under the doctrine of equivalents.

Interestingly enough, although the Federal Circuit in *Lemelson* cited *Hughes Aircraft* for other propositions, it made no mention of the requirement in *Hughes Aircraft* that the invention be viewed as a whole.

---

in light of prior art. The appellate court reversed this finding; however, a review of the cited prior art indicates that the original patent cannot be construed as more than a minor advancement in the technological area of television rental systems.)

96. *See* George v. Honda Motor Co., 802 F.2d 432 (Fed. Cir. 1986) (Patent owner claimed that defendant infringed upon his patented internal combustion engine that contained a water-cooled cylinder and an air-cooled cylinder head. The accused device had a water-cooled cylinder head with a water jacket which covered the head. The Federal Circuit held that there was no infringement under the doctrine of equivalents because the air cooled cylinder head was an essential part of the claimed invention.) George is similar to ACS Hospitals in that the claims of the original patents are both comprised of three elements and the absence of one of these elements in the accused device precluded a finding of infringement. In both cases, the Federal Circuit found that the missing element was highly important to the patented combination, and thus accorded a narrower range of equivalents.

97. 752 F.2d 1538 (Fed. Cir. 1985).

98. *Id.*

99. 531 F.2d 547 (Ct. Cl. 1976); *see supra* text accompanying notes 58-64.

100. 752 F.2d at 1551 (Fed. Cir. 1985).

However, a review of the court's infringement analysis reveals that it impliedly adopted the invention as a whole approach when determining the range of equivalents to be accorded the manipulation means.[101] Both *ACS Hospital* and *Lemelson* demonstrate the implicit integration of the invention as a whole approach in cases where the apparent application of the "all elements" rule would lead the unwary to believe that only the element-by-element analysis was used.

D. Texas Instruments: Extension of the Invention as a Whole Standard to Preclude Infringement Even When All Elements Are Substantially Similar

In a decision rendered in 1986 and written by Judge Newman,[102] *Texas Instruments, Inc. v. International Trade Commission*,[103] the Federal Circuit applied the invention as a whole approach, not only in the context of infringement under the doctrine of equivalents, but also in the issue of literal infringement.[104] The Federal Circuit's opinion in *Texas Instruments* has been criticized by the patent community for its interpretation of the invention as a whole standard that was ambiguously set forth in *Hughes Aircraft*.[105] Rather than using the invention as a whole requirement as an admonition to view each element in the context of the entire claim, the Federal Circuit instead used it as authority for comparing the accused device with the claimed invention as a whole.[106] It was this novel use of the invention as a whole requirement in *Texas Instruments* that has caused much of the current confusion over the standard of applicability for the doctrine of equivalents.

In *Texas Instruments*, the Federal Circuit upheld a decision of the United States International Trade Commission that the means-plus function claims of a pioneer patent for a portable electronic calculator were

---

101. *Id.*

102. Judge Newman, as shown by her dissenting commentary in *Pennwalt*, is one of the strongest proponents of the invention as a whole standard to the exclusion of the element- by-element analysis.

103. 805 F.2d 1558 (Fed. Cir. 1986), *reh'g denied* 846 F.2d 1369 (Fed. Cir. 1988), *reh'g denied*, (en banc) 7 U.S.P.Q.2d (BNA) 1414 (Fed. Cir. 1988).

104. *Id.*

105. *See* Nieman, *supra* note 50, at 154-5.

> The *Texas Instruments* opinion was greeted by a firestorm of criticism from the patent bar . . . . The [American Intellectual Property Law Association] interpreted the *Texas Instruments* opinion as signalling a rejection of an element-by-element infringement analysis under Section 112 (6), or the doctrine of equivalents in favor of viewing the claimed invention 'as a whole' divorced from adherence to the claim language.

Id.

106. 805 F.2d at 1571.

not infringed by imported calculators.[107] First, the Federal Circuit found that the functional claims of the original patent were not literally infringed by the accused device.[108] The court reviewed each of the functional clauses in the original patent[109]— the input means, the electronic means, and the display means — and concluded that although the Commission's finding of nonequivalence was not supported by the evidence when each element is considered separately, "the accused devices do not infringe properly construed claims when the invention and the accused devices are viewed as a whole."[110] This is similar to the reverse doctrine of equivalents[111] where the court will apply the doctrine of equivalents to defeat a finding of infringement where the accused device falls within the literal language of the claims but is so far changed from the original invention that the patent owner of the original claims has no right to the accused device.[112] The Federal Circuit's use of the "invention as a

---

107. *Id.*

108. *Id.*

109. *Id.* at 1570.

> To summarize the totality of changes: The input means in the '921 patent is a keyboard encoder that operates through conductive strips under the keys, whereas in the accused devices it is a scanning matrix encoder. The electronic means in the '921 patent is an integrated semiconductor array based on bipolar semiconductor technology; the accused devices use metal oxide semiconductors and embody significant advances in chip design and integrated circuitry. The display means in the '921 patent is a thermal printer, whereas the accused calculators use liquid crystal displays.

Id.

110. *Id.* at 1564, 1570.

> [The] accumulated differences distinguish the accused calculators from that contemplated in the '921 patent and transcend a fair range of equivalents of the '921 invention. Each individual difference, standing alone, could conceivably lead to a different result, by application of this court's precedent. It is to the invention as a whole to which this same precedent directs our analysis.

Id.

111. *See supra* note 41.

112. *See supra* note 41. Remember that equivalents of means-plus function clauses are deemed to fall within the literal language of the claims, and comprise an analysis that is separate from the determination of the functional equivalence in the doctrine of equivalents. *See supra* text accompanying notes 11-14. In her denial of petition for rehearing, Judge Newman later refuted suggestions by the amicus that the reverse doctrine of equivalents was applicable. 846 F.2d 1369, 1371 (Fed. Cir. 1988). She explained that the court had held that there was no literal infringement even though the functions of the accused devices and the original patent were equivalent because the structures performing those functions were non-equivalent. *Id.* Therefore, Judge Newman flatly concluded that without the prerequisite of literal infringement, the reverse doctrine of equivalents cannot be invoked. *Id.*

This conclusion ignores the underlying meaning of the reverse doctrine of equivalents— that literal infringement is precluded when the accused device falls within the literal

whole" language in its literal infringement analysis[113] is merely a somewhat careless substitution for the usual reverse doctrine of equivalents language that the accused device does not lie within the scope of the claims of the original patent because "it performs the same or similar function in a substantially different way."[114]

Second, in the Federal Circuit's analysis of whether the accused device infringed the original patent under the doctrine of equivalents, the Federal Circuit states that the proper analysis consists of a comparison of the accused device with the claimed invention as a whole.[115] To follow this usage of the invention as a whole requirement would in effect, exclude any element-by-element analysis. Under the *Texas Instruments* approach to the invention as a whole requirement, an accused device could be found to infringe a patented claim even if it failed to contain an element of the claim. This potential outcome is in clear derogation of the "all elements" rule which states that a finding of infringement is precluded against all accused devices that do not contain an essential

---

language of the claims, but performs the same or similar function in a different way. In holding that the same functions were performed by the accused device but that the structures performing those functions were different, the court basically held that the same functions were performed in a different way. This is the essence of the logic behind the reverse doctrine of equivalents.

Even Judge Davis, who joined in both the original opinion and part of the denial of petition for rehearing, stated that in deciding the original case, he applied an analogy to the reverse doctrine of equivallents. *Id.* at 1372.

> I agree with the result and join all of Judge Newman's opinion except the last portion, headed 'The Reverse Doctrine of Equivalents.' My view is that both the original opinion in this case (which I joined) and the opinion on rehearing (the greater part of which I have joined) reflect principles comparable to (though distinct from) the reverse doctrine of equivalents . . . . I have considered it proper to apply here an analogue or parallel of the reverse doctrine of equivalents—though not that doctrine in and of itself.

Id. Therefore, the literal infringement analysis in *Texas Instruments* may fairly be read as embodying the spirit of the reverse doctrine of equivalents principles. *See also* Nieman, *supra* note 50, at 160. ("*Texas Instruments* must be deemed to be a *sub silentio* 'reverse doctrine' case in light of the *in banc* decision in *Pennwalt.*")

113. *Texas Instruments*, 805 F.2d at 1568-71.

114. *Graver Tank*, 339 U.S. 605, 608-9 (1950). *See also* Mead Digital Systems, Inc. v. A. B. Dick Co., 723 F.2d 455, 462 (6th Cir. 1983).

> [T]he doctrine of equivalents is a two-edged sword: 'where a device is so far changed in principle from a patented article that it performs the same or similar function in a substantially different way, but nevertheless falls within the literal words of the claim, the doctrine of equivalents may be used to restrict the claim and defeat the patentee's action for infringement.' The doctrine thus may be applied in favor of as well as against a patentee.

Id. (citations omitted.)

115. 805 F.2d at 1571.

element of the patented claim.[116] This inconsistency with prior case law renders completely unviable any analysis of infringement under the doctrine of equivalents that consists solely of a comparison of the accused device and the patented invention as a whole.

Despite the court's statement that its analysis of infringement under the doctrine of equivalents depended solely on a comparison of the accused device with the invention as a whole, the court impliedly used an element-by-element analysis. In the section of the opinion discussing the doctrine of equivalents, the court referred to its reasoning on the issue of literal infringement.[117] A review of the court's discussion of the issue of literal infringement reveals that the court examined each claim in the original patent to determine if it or its equivalent was present in the accused device.[118] This type of analysis is indicative of an element-by-element analysis. In sum, the court addressed the issue of literal infringement using an element-by-element analysis, then incorporated this analysis by reference in its determination of infringement under the doctrine of equivalents. The court has, in actuality, not relied solely on an invention as a whole analysis, but has relied on an element-by-element analysis as well.[119] Therefore, *Texas Instruments* may be viewed as standing for the following proposition: in an infringement analysis under the doctrine of equivalents, after determining by an element-by-element analysis that the accused device contains an equivalent for each element in the patented claim, a court may still deliver a finding of non-infringement if it finds that a comparison of the accused device with the claimed invention as a whole reveals that the totality of the differences renders the two nonequivalent.[120]

E. Perkin-Elmer: Federal Circuit Explicitly Integrates Both Invention as a Whole and Element-by-Element Standards in the Same Case

Finally, a few months before the *Pennwalt* case was decided, the Federal Circuit in *Perkin-Elmer Corp. v. Westinghouse Electric Co.*[121]

---

116. *See supra* text accompanying notes 59-62, 87-89.

117. 805 F.2d at 1572.

> In this case, the determination turns on the totality of change in the accused devices from that described in the '921 specification. For the reasons discussed in part A [literal infringement], the extensive technological advances in all of the claimed functions support the ALJ's finding that the accused devices are not equivalent to the claimed invention, applying the criteria of *Graver Tank*.

Id.

118. *Id.* at 1561-71.

119. *Id.*

120. *Id.*

121. 822 F.2d 1528 (Fed. Cir. 1987).

expressly sanctioned the use of an integration of both the invention as a whole standard and the element-by-element standard in infringement analyses.[122] Chief Judge Markey explained that his adoption of the invention as a whole approach in *Hughes Aircraft* did not reject the element-by-element analysis.[123]

> In *Hughes Aircraft Co. v. United States*, this court noted that it was legal error not to 'apply the doctrine of equivalents to the claimed invention as a whole.' That statement dealt with an infringement inquiry implicating an entire claim, as distinguished from a section 112 (6) inquiry implicating only a "means plus function" limitation of a claim. That statement also was a recognition that, in applying the doctrine of equivalents, each limitation must be viewed in the context of the entire claim. The statement should not be interpreted as sanctioning the treatment of claim limitations as insignificant or immaterial in determining infringement.[124]

Chief Judge Markey then reiterated the now familiar "all elements" rule, adding that in order to be considered a substantial equivalent for purposes of the rule, the element substituted in the accused device for the element set forth in the claim must not substantially change the manner in which the original invention functions.[125]

In clarifying the invention as a whole language as it was presented originally in *Hughes Aircraft*, the Federal Circuit has affirmed the usage of both the invention as a whole standard and the element-by-element standard into a single framework for analyzing issues of infringement under the doctrine of equivalents. According to *Perkin-Elmer*, these two distinct standards intertwine in the following way. The first tier of inquiry is whether the plaintiff can show the presence of every element of his claim in the accused device.[126] If certain elements in his claim are missing from the accused device, the inquiry now becomes whether he can show the presence of substantial equivalents of those elements.[127] To determine the range of equivalents to be accorded the disputed elements, each element must be viewed in the context of the entire claim.[128]

In *Perkin-Elmer*, the patent in question is a resonant coupler for an electrodeless discharge lamp.[129] The original invention uses an au-

---

122. *Id.*
123. *Id.* at 1532-33.
124. *Id.* (citations omitted).
125. *Id.*
126. *Id.*
127. *Id.*
128. *Id.*
129. *Id.* at 1322-33.

totransformer-type tap coupling, while the accused device uses a transformer-type loop coupling.[130] In these two types of couplings, the connecting points between the helical coil and the power source and the coupling arrangements are different.[131] The majority held that these differences establish that the two types of couplings do not work in substantially the same way, and thus, there is no infringement under the doctrine of equivalents.[132] In a dissenting opinion, Judge Newman argued that the accused device does infringe under the doctrine of equivalents, based on a finding that the two types of couplings work in substantially the same way.[133]

Judge Newman contended that the majority failed to follow case precedent, and misapplied the invention as a whole standard.[134] This criticism of the majority's opinion is not unexpected when considering that the majority impliedly disapproved of the application of the invention as a whole standard to means-plus function inquiries, such as was done in *Texas Instruments*.[135] Although *Perkin-Elmer* clarifies the application of the invention as a whole standard in *Hughes Aircraft*,[136] it still leaves unresolved the applicability of the invention as a whole standard in *Texas Instruments* to means-plus function clauses in issues of literal infringement. It also does not address the fate of the holding in *Texas Instruments* that the invention as a whole standard can be applied to prevent a finding of infringement even where the elements, when viewed separately, are substantially equivalent.

F. District Court Opinions

The ambiguities in the Federal Circuit's treatment of the standard of applicability for the doctrine of equivalents have been echoed in district court opinions.[137] A review of recent opinions prior to *Pennwalt* reveals that most district courts are unclear as to what the standard of analysis for the doctrine of equivalents is, thereby explicitly adopting one standard while implicitly integrating it with the reasoning of the other.[138] Most opinions explicitly adopt an element-by-element analysis.[139]

---

130. *Id.*
131. *Id.* at 1323.
132. *Id.* at 1326-27.
133. *Id.* at 1329 (Newman, J., dissenting).
134. *Id.* at 1327 (Newman, J., dissenting).
135. 805 F.2d 1558 (Fed. Cir. 1986).
136. 717 F.2d 1351 (Fed. Cir. 1983).
137. *See infra* notes 139-40.
138. *Id.*
139. Amicus Inc. v. American Cable Co., 660 F. Supp. 161 (E.D. La. 1987); Weber v. Ford Motor Co., 664 F. Supp. 631 (D. Mass. 1987); Westnofa USA Inc. v. Whole

Only a few opinions explicitly apply the invention as a whole approach.[140]

III. Pennwalt Corp. v. Durand-Wayland, Inc.

In *Pennwalt Corp. v. Durand-Wayland, Inc.*,[141] Pennwalt sued Durand-Wayland for alleged infringement of Pennwalt's patented fruit sorter.[142] The majority opinion of the Federal Circuit written by Judge Bissell, and joined by Judges Friedman, Rich, Davis, Nies, Archer, and Chief Judge Markey (author of the *Hughes Aircraft* and *Perkin-Elmer* decisions), affirmed the trial court's finding that Durand-Wayland's fruit sorter did not infringe Pennwalt's claims either literally or under the doctrine of equivalents.[143] Instrumental to the majority's decision was the following quote from *Lemelson*:

> [I]n applying the doctrine of equivalents, *each limitation must be viewed in the context of the entire claim* . . . . It is . . . well settled that each element of a claim is material and essential, and that in order for a court to find infringement, the plaintiff must show the presence of *every element or its substantial equivalent* in the accused device.[144]

In relying upon the above quote from *Lemelson*, the majority adopted an infringement analysis that combines both the invention as a whole approach and the element-by-element standard.

The majority rejected Pennwalt's arguments that the accused devices performed substantially similar functions in substantially the same way to achieve the same result, in this case, the sorting of fruit.[145] In doing so, the majority relied upon the trial court's findings that the accused device was missing certain functions of the patented fruit sorter and that the functions which were performed by the accused device were

---

Life Co., 3 U.S.P.Q.2d 1352 (D. Mass. 1987); Intra Corp. v. Hamar Laser, 662 F. Supp. 1420 (E.D. Mich. 1987), *cert. denied*, 109 S. Ct. 1746 (1989); Baker Oil Tools v. TRW Inc., 673 F. Supp. 1061 (N.D. Okla. 1987).

140. Moleculon Research v. CBS, 666 F. Supp. 661 (D. Del. 1987), *rev'd*, 872 F.2d 407 (Fed. Cir. 1989); Key Manufacturing Group Inc. v. Microdot, 679 F. Supp. 648 (E.D. Mich. 1987), *aff'd in part, vacated in part*, 854 F.2d 1328 (Fed. Cir. 1988). *But cf.* Intra Corp. v. Hamar Laser, *supra* note 139; Simmons Fastener Corp. v. Illinois Tool, 663 F. Supp. 697 (N.D.N.Y. 1987).

141. 833 F.2d 931 (Fed. Cir. 1987), *cert. denied*, 108 S. Ct. 1226 (1988), *cert. denied*, 108 S. Ct. 1474 (1988).

142. *Id.*

143. *Id.*

144. *Id.* at 935 (quoting Lemelson v. United States, 752 F.2d 1538, 1551 (Fed. Cir. 1985)) (emphasis added).

145. *Id.* at 935.

substantially different.[146] "[T]he district court correctly relied upon an element-by-element comparison to conclude that there was no infringement under the doctrine of equivalents, because the accused devices did not perform substantially the same functions as the Pennwalt invention."[147] The language of the majority opinion makes it clear that the backbone of an infringement analysis under the doctrine of equivalents is the element-by-element standard.

However, the majority also used the invention as a whole approach implicitly when it viewed each element in the context of the entire claim.[148] For example, the majority agreed with the trial court's finding that the range of equivalents for the position-indicating means is narrow for two reasons.[149] The first reason is because the entire invention was not a pioneer, but only "an improvement in a crowded field," thereby permitting only a narrow range of equivalents.[150] The second limitation upon the range of equivalents results from the prosecution history which indicated that the addition of the position-indicating means was crucial to the invention's patentability.[151] Therefore, the inventor is estopped from obtaining a range of equivalents that expands upon the specific functional limitations that were necessary in order to secure his patent.[152]

The majority in *Pennwalt* also resolved an ambiguity that arose in case precedent after the *Texas Instruments* opinion: the application of a means-plus function equivalents analysis to the doctrine of equivalents.[153] Means-plus function equivalence in a literal infringement inquiry requires exact equivalence.[154] "If the required function is not performed exactly in the accused device," then means-plus equivalents for the purposes of literal infringement are not involved.[155] The majority in *Pennwalt* then proceeded to state that means-plus equivalents inquiries for literal infringement play no role in determining whether the accused device performs an equivalent function for purposes of the tripartite *Graver Tank* test under the doctrine of equivalents.[156]

---

146. *Id.*
147. *Id.*
148. *Id.* at 934-39.
149. *Id.* at 937.
150. *Id.* ("The claims are 'broad' with respect to what type of product can be sorted, i.e., 'items' and, thus, sorters of all types fall within the relevant prior art. The claims are *narrow*, however, with respect to how the claimed sorter operates.") (emphasis in original).
151. *Id.*
152. *Id.* at 938.
153. *See supra* text accompanying notes 117-18.
154. 833 F.2d at 934.
155. *Id.*
156. *Id.*

The dissenting opinion written by Judge Bennett and joined by Judges Cowen, Smith and Newman (the author of the *Texas Instruments* opinion) expressed concern that an element-by-element comparison would undermine the concept of the doctrine of equivalents.[157] Such an analysis, the dissent feared, would be the equivalent of literal infringement and would be "so unduly restrictive and inflexible as to end its usefulness as a judicial doctrine."[158]

> In fact, the majority's current analysis under the doctrine of equivalents amounts to nothing more than the search of 'obvious and exact equivalents' that this court denounced in *Hughes*. . . . In my view, the majority's adoption of a rigid element-by-element analysis for the doctrine of equivalents skews the balance well toward a slavish and formalistic adherence to the words of a claim to the detriment of both the patent owner and the public.[159]

The dissent protests that in requiring an element-by-element analysis, the majority overruled *Hughes Aircraft* without even addressing the case in its opinion.[160]

The basic principle underlying the dissent's argument, that the element-by-element standard and the invention as a whole standard are mutually exclusive, is incorrect. In stating that the trial court correctly applied an element-by-element analysis, the majority was merely reiterating the well-established "all elements" rule.[161] By approving the element-by-element standard, the majority did not reject the invention as a whole approach *sub silentio*. The majority implicitly incorporated the invention as a whole approach in its determination of the range of equivalents to be accorded the disputed elements.[162] The dissent failed to recognize the majority's usage of the invention as a whole restriction simply because the dissent never started with an accurate picture of invention as a whole restriction. The dissent views the invention as a whole standard which was first set forth in *Hughes Aircraft* to require a comparison of the accused device and the claimed invention as a whole to determine if the tripartite test of *Graver Tank* is satisfied.[163] As *Perkin-Elmer* indicates, this is an inaccurate view of the invention as a whole standard.[164] The invention as a whole standard is merely a re-

157. *Id.* at 939-40 (Bennett, J., dissenting).
158. *Id.* at 940.
159. *Id.* at 942, 948.
160. *Id.* at 941.
161. *Id.* at 949-52 (Nies, J., additional views).
162. *See supra* text accompanying notes 148-152.
163. 833 F.2d at 948.
164. 822 F.2d 1528 (Fed. Cir. 1987).

quirement that each element be viewed in the context of the entire claim.[165] Its integration into the element-by-element analysis is not only possible but required.

Furthermore, the dissent's view of the invention as a whole restriction invites short-sighted policy considerations. Although the doctrine of equivalents was created to prevent inequity, courts are not free to disregard the statutory requirement that patent claims set forth the boundaries of what the applicant regards as his invention.[166] If courts are permitted to find infringement even where the accused device lacks an element of the patented claim, then the competitors' due process rights are violated because they no longer have notice of what constitutes a technological advance sufficient to avoid infringement.[167]

Nor will the dissent's view of the invention as a whole standard aid patent owners because they will likewise be uncertain where the courts will draw the line on the issue of infringement. An approach that allows a court to find infringement based upon its view of what the invention encompasses rather than what the elements in the claim set forth may even encourage sloppy drafting of claims because the vaguer the claim, the more leeway the court has in interpreting what the claim encompasses. One of the fundamental goals of patent law is to promote consistency and predictability in case decisions so that patent owners and competitors alike can know what is currently protected by patents and what is not. An infringement analysis that leaves the determination of what the patented claim encompasses to the subjective views of the court will only serve to add uncertainty to case law.

IV. Framework for Analysis of Infringement Issues After *Pennwalt*

The *Pennwalt* decision, together with *Perkin-Elmer*, greatly clarified the standard of analysis for infringement issues under the doctrine of equivalents. Both cases show that the correct method of analysis is an integration of the element-by-element standard and the invention as a whole approach. The following algorithm suggests the framework which the Federal Circuit will apply in future cases involving infringement issues.[168]

---

165. *Id.*

166. 35 U.S.C. 112 (1982).

167. Pennwalt Corp. v. Durand-Wayland, Inc., 833 F.2d 931, 954 (Fed. Cir. 1987).

168. A recent case, May v. Carriage, adopts the following methodology for infringement analyses after *Pennwalt* which was set forth in the Nieman article, *see supra*

(1) Claim Construction—Determine scope of the claims by examining:

(a) the language of the claims themselves,

(b) the specification,

(c) the drawings, and

(d) the file wrapper.[169] File wrapper estoppel will prevent the inventor from claiming that subject matter which he surrendered during the prosecution of his patent application lies within the scope of his claims.[170]

(2) Literal Infringement

(a) After ascertaining the scope of the claims, compare each element of the claims to the accused device. If all the elements in the claim read upon the accused device, then there is literal infringement of the claims.[171]

(1) Structural elements must be copied in order to be literally infringed.[172]

(2) Means-plus functional elements are deemed to be literally infringed if the accused device contains equivalents of the means clauses.[173]

---

note 50. 688 F. Supp. 408 (N.D. Ind. 1988).

> Before *Pennwalt*, it was unclear whether, from an analytical viewpoint, the doctrine of equivalents was to be applied on an 'element-by-element basis' or to the invention 'as a whole.' Under *Pennwalt*, a doctrine of equivalents analysis should proceed as follows:
>
> 1. Determine whether the accused device achieves substantially the same result as the claimed invention.
>
> 2. Determine whether the accused device performs substantially the same work as the claimed invention.
>
> 3. Determine whether the accused device operates in substantially the same manner as the claimed invention. In so doing, compare each element of the claim with the accused device to determine whether the accused device contains each element or its substantial equivalent. *Pennwalt*, therefore, has adopted the element-by-element analysis.

688 F. Supp. at 422-23.

The above methodology is based on the erroneous view that the element-by-element standard and the invention as a whole standard are mutually exclusive. Because this methodology fails to incorporate any consideration of the invention as a whole standard, it fails to state the entire algorithm necessary for an infringement analysis under the doctrine of equivalents. Moreover, the issue of infringement in May v. Carriage is dicta because the court held that the patent was invalid for obviousness and thus cannot be infringed. *Id.* at 422.

169. *See supra* note 31.

170. *See supra* notes 34-39 and accompanying text.

171. Graver Tank & Mfg. Co. v. Linde Air Products Co., 339 U.S. 605, 608-09 (1950).

172. *See supra* notes 12-14 and accompanying text.

173. *See supra* notes 12-14 and accompanying text.

To determine equivalence, the court will compare "the accused structure with the disclosed structure, and must find equivalent structure as well as identity of the claimed function for that structure."[174]

(b) Before reaching a conclusion of literal infringement, ensure that the defense of the reverse doctrine of equivalents does not apply. If the accused device falls within the literal language of the claim but is "so far changed in principle from the patented article that it performs the same or similar function in a substantially different way," then a finding of literal infringement is precluded by the reverse doctrine of equivalents.[175]

(c) Only if there is no literal infringement of the claims should the court then proceed to the infringement analysis under the doctrine of equivalents.

(3) Infringement under the Doctrine of Equivalents

(a) View the claimed invention as a whole to determine the range of equivalents to be accorded the entire invention on a continuum from a broad range of equivalents for a pioneer invention to a minimal range of equivalents for a small advancement.[176]

(b) Compare the claimed invention to the accused device to determine whether they accomplish substantially the same results. This is the first part of the tripartite *Graver Tank* test, and unless the two inventions perform substantially the same end results, there is no infringement under the doctrine of equivalents.[177]

(c) Compare each element of the claimed invention to the accused device to determine whether they accomplish substantially the same functions in substantially the same manner. This is a combination of the second and third part of the tripartite *Graver Tank* test, and unless it is satisfied, there is no infringement under the doctrine of equivalents.[178]

(1) All elements of the claimed invention are material and essential.[179] Therefore, omission of any element or its equivalents by the accused device precludes a finding of infringement.[180] To determine whether the disputed element has an equivalent in the accused device:

---

174. *Pennwalt*, 833 F.2d at 934.

175. *See supra* note 114.

176. *See supra* notes 23-6 and accompanying text. *See also Corning Class Works v. Sumitomo Electric, U.S.A., Inc.*, 868 F.2d 1251, 1260 (Fed. Cir. 1989) ("[T]he court made a subsidiary analysis comparable to the overall function/way/result analysis mandated for determining infringement of the claim under the doctrine of equivalents.")

177. *See supra* text accompanying notes 5-8.

178. *See* Graver Tank & Mfg. Co. v. Linde Air Products, 339 U.S. 605 (1950).

179. *See supra* notes 87-90 and accompanying text.

180. *See supra* notes 87-90 and accompanying text.

(a) Determine the range of equivalents of each element by viewing it in the context of the entire claim.[181]

(b) Accord elements which are the major features of the claimed invention a narrower range of equivalents, while elements that are minor features should be accorded a broader range of equivalents.[182] As a general proposition, note that when the total number of elements in a claim diminishes, each element becomes more important to the claim as a whole, and should be accorded a narrower range of equivalents.[183]

(c) One-to-one equivalence between the corresponding components of the accused device and the original patent is not necessary for a finding of elemental equivalence.[184] So long as an equivalent is

---

181. *See supra* text accompanying notes 124-28.

182. R. Ellis, Patent Claims, 33 (1949) ("An invention, however, usually has minor as well as major features. The doctrine of equivalents may be invoked as to either or both the minor and major features. . . . The less important the minor feature is relatively to the main feature, the wider the range of equivalents as applied to the minor feature. This is due to the fact that a subsidiary feature of an invention can be radically changed with less effect on the invention as a whole than a relatively minor change in the main feature of the invention.")

183. *See supra* note 96 and accompanying text.

184. A post-*Pennwalt* decision, *Corning-Glass Works v. Sumitomo Electric, U.S.A., Inc.*, demonstrates the Federal Circuit's application of the doctrine of equivalents to reach a finding of infringement. 868 F.2d 1251 (Fed. Cir. 1989). In *Corning-Glass*, the inventions under scrutiny relate to the structure and method for making optical waveguide fibers. *Id.* at 1253-55. The original patent disclosed an optical fiber in which a certain dopant in the core increased the refractive index of the core (a "positive dopant"), and caused a specific refractive index differential between the core and the cladding. *Id.* It was this specific refractive index differential which, together with a selected core diameter, allowed light to be transmitted for far greater distances than was previously possible with optical fibers. *Id.* at 1254.

The accused device was similar to the original patent except that instead of adding a dopant to the core to increase the refractive index of the core, it contemplated adding a dopant to the cladding to decrease the refractive index of the cladding (a "negative dopant"). *Id.* at 1259. The accused infringer argued that because it did not substitute an element which increased the refractive index of the core, an element required by the original patent is missing. *Id.* Therefore, the accused infringer concluded that because the "all elements" rule requires the presence of every element of the original patent or its equivalent in the accused device before infringement can be found, it did not infringe the original patent. *Id.*

Ruling that one-to-one equivalence between corresponding components is not mandatory, the Federal Circuit held that the element of a positive dopant in the core is not entirely missing from the accused device. *Id.*

> 'Element' may be used to mean a single limitation, but it has also been used to mean a series of limitations which, taken together, make up a component of the claimed invention. In the All Elements rule, 'element' is used in the sense of a limitation of a claim. . . . *An equivalent must be found for every limitation*

found for every limitation of the claim somewhere in the accused device, the "all elements" rule is satisfied.[185]

(2) A finding of infringement will not be precluded merely because the accused device contains elements which the patented claim does not have.[186]

(d) If the above analysis for the tripartite *Graver Tank* test for equivalence is satisfied, then the accused device infringes the claimed invention under the doctrine of equivalents. However, the effect of *Texas Instruments* on this outcome is somewhat uncertain at this date. Courts may follow *Texas Instruments* and hold that there is no infringement where a comparison of the accused device with the claimed invention as a whole reveals that the totality of the differences renders the two nonequivalent.[187]

---

of the claim somewhere in an accused device, but not necessarily in a corresponding component, although that is generally the case.

Id. (citations omitted) (emphasis added). The court then approved the district court's analysis which compared the substituted limitation with the limitation in the context of the invention through the use of the tripartite *Graver Tank* test. *Id.* at 1260. Agreeing that the negative dopant in the cladding "performs substantially the same function in substantially the same way as the use of a positive dopant in the core to produce the same result of creating the refractive index differential between the core and cladding of the fiber which is necessary for the fiber to function as an optical waveguide," the Federal Circuit affirmed the district court's finding of the infringement under the doctrine of equivalents. *Id.*

A recent article suggests that *Corning-Glass* stands for the proposition that an accused device may be held to be the equivalent of the original patent even if the substitute element performs the opposite function of the claim element. Player, *Elemental Equivalence: Interpreting "Substantially the same way" in Pennwalt after Corning Glass*, 71 J. PAT. OFF. SOC'Y 546 (1989). Because the negative dopant in the accused device lowered the refractive index of the cladding while the positive dopant in the original patent increased the refractive index of the core, the article writer reasoned that for purposes of the doctrine of equivalents, the two performed opposite functions. *Id.* at 553. This analysis, however, ignores the admonition that the claims be construed in the context of the entire invention. In *Corning-Glass*, the lowering of the refractive index of the cladding performs the same function as the raising of the refractive index of the core: the creation of a refractive index differential which traps light within the optical fibers. To require the plaintiff to show infringement under the doctrine of equivalents by demanding the presence of a positive dopant in the core is precisely the type of redundant literal infringement inquiry which the Federal Circuit has consistently struck down. *See Hughes Aircraft*, 717 F.2d 1351.

185. *Id.*

186. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 1057 (Fed. Cir. 1988) ("Adding features to an accused device will not result in noninfringement if all the limitations in the claims, or equivalents thereof, are present in the accused device."), *cert. denied*, 109 S. Ct. 75 (1988).

187. *See supra* text accompanying note 120.

V. Conclusion

The standard for an infringement analysis under the doctrine of equivalents is a combination of both the element-by-element standard and the invention as a whole approach. This integrated standard not only reconciles case precedent on this controversy, but more importantly, best fulfills the purpose of the doctrine of equivalents. Case law addressing the doctrine of equivalents reflects the tension between the competing policies of expanding the meaning of the claims in a patent and thereby protecting the patent owner, and giving competitors fair notice of what the patent owner believes to be the boundaries of his invention. Adoption of the misinterpretation of the invention as a whole restriction set forth by the court in *Texas Instruments* would skew the analysis in favor of the patent owner and deprive his competitors of fair notice of what constitutes protected material. Likewise, adoption of the element-by-element analysis without the pragmatic requirement that each element be viewed in the context of the invention as a whole would lead to inequitable results because minor deviances could be viewed disproportionately and vice versa. An analysis that utilizes both the element-by-element standard and the invention as a whole requirement supplies both consistency and fairness—two sought-after qualities in any area of the law.

Emily Lau

The Standard of Proof in Civil RICO Actions for Treble Damages: Why the Clear and Convincing Standard Should Apply

I. Introduction

The Racketeer Influenced and Corrupt Organizations Act (RICO)[1] was passed by Congress in 1970 as a part of the Organized Crime Control Act of 1970.[2] The RICO legislation provides civil remedies[3] as well as criminal remedies.[4] Civil actions may be brought by the Attorney General of the United States[5] or by any person harmed in his person or business by a violation of the predicate criminal acts.[6]

The substantive requirements of a civil RICO action are identical to those for a criminal proceeding.[7] In other words, the plaintiff in a civil action will only be entitled to relief once the elements of a criminal violation of the statute have been proven. A prior conviction of a criminal violation of the RICO statute is not necessary to bring a civil action,[8] although it is necessary to prove the predicate elements of a criminal violation.[9]

A logical issue in this path of reasoning is a determination of the proper standard of proof for the predicate criminal acts in a civil RICO action. Since the predicate acts are necessarily criminal and since it is not necessary to show a prior conviction on these acts to prevail, should the plaintiff be held to a higher standard of proof than the ordinary civil standard of "by a preponderance of the evidence"?

The Supreme Court has stated in dicta that the standard of proof to be applied to the predicate criminal acts is lower than the criminal standard of "beyond a reasonable doubt."[10] However, the Court specifically left unanswered the question of what the proper standard of proof should be.[11] Several of the circuit courts have stated in dicta that

---

1. 18 U.S.C. 1961-1968 (1982).
2. Pub. L. No. 91-452, 84 Stat. 922 (1970).
3. 18 U.S.C. 1964 (1982).
4. 18 U.S.C. 1963 (1982).
5. 18 U.S.C. 1964(b) (1982).
6. 18 U.S.C. 1964(c) (1982).
7. 18 U.S.C. 1964 (1982).
8. Sedima, S.P.R.L. v. Imrex Co., 473 U.S. 479, 491 (1985).
9. 18 U.S.C. 1963 (1982).
10. *Sedima*, 473 U.S. at 491.
11. *Id.*

the proper standard in a civil RICO action is the ordinary civil action standard of preponderance of the evidence.[12]

Recently, the Seventh Circuit was the first appellate court to address the standard of proof question when it was directly in issue.[13] The Seventh Circuit held that the proper standard of proof in a civil RICO action is the ordinary civil "preponderance" standard.[14]

There remain several circuits which have not addressed the standard of proof issue.[15] The time is ripe to closely analyze the issue and decide the appropriate standard of proof based on a complete consideration of all of the relevant factors. This Note will advocate that the proper standard of proof is the higher standard of "clear and convincing evidence," taking a position opposite to the courts which have already addressed the issue.

A. The RICO Statute

The RICO statute is a powerful tool to use against those who violate specified criminal acts. The statute makes it unlawful to invest income derived from a pattern of racketeering activity in any enterprise affecting interstate commerce,[16] acquire or maintain an interest in such an enterprise

---

12. Wilcox v. First Interstate Bank, 815 F.2d 522, 531 (9th Cir. 1987); Cullen v. Margiotta, 811 F.2d 698, 731 (2d Cir.), *cert. denied*, 107 S. Ct. 3266 (1987); Armco Indus. Credit Corp. v. SLT Warehouse Co., 782 F.2d 475, 480-481 (5th Cir. 1986); U.S. v. Local 560 of Int'l Bhd. of Teamsters, 780 F.2d 267, 279-280 n.12 (3d Cir. 1985), *cert. denied*, 476 U.S. 1140 (1986).

13. Liquid Air Corp. v. Rogers, 834 F.2d 1297, 1302-1303 (7th Cir 1987)(U.S. App. pending.).

14. *Id.*

15. The first, fourth, sixth, eighth and tenth circuits have not addressed the standard of proof issue in civil RICO.

16. 18 U.S.C. 1962(a) (1982).

> (a) It shall be unlawful for any person who has received any income derived, directly or indirectly, from a pattern of racketeering activity or through the collection of an unlawful debt in which such person has participated as as a principal within the meaning of section 2, title 18, United States Code, to use or invest, directly or indirectly, any part of such income, or the proceeds of such income, in acquisition of any interest in, or the establishment or operation of, any enterprise which is engaged in, or the activities of which affect, interstate or foreign commerce. A purchase of securities on the open market for the purposes of investment, and without the intention of controlling or participating in the control of the issuer, or of assisting another to do so, shall not be unlawful under this subsection if the securities of the issuer held by the purchaser, the members of his immediate family, and his or their accomplices in any pattern or racketeering activity or the collection of an unlawful debt after such purchase do not amount in the aggregate to one percent of the outstanding securities of any one class, and do not confer, either in law or in fact, the power to elect one or more directors of the issuer.

through a pattern of racketeering activity,[17] conduct or participate in the affairs of such enterprise through a pattern of racketeering activity,[18] or conspire to do any of these activities.[19] "Racketeering activity" includes a range of offenses from kidnapping to mail and interstate wire fraud.[20] To come under the statute, at least two of the predicate offenses have to be committed within ten years of each other and must constitute a "pattern."[21] Additionally, the offenses must be committed within an

---

17. 18 U.S.C. 1962(b) (1982).

(b) It shall be unlawful for any person through a pattern of racketeering activity or through collection of an unlawful debt to acquire or maintain, directly or indirectly, any interest or control of any enterprise which is engaged in, or the activities of which affect, interstate or foreign commerce.

18. 18 U.S.C. 1962(c) (1982).

(c) It shall be unlawful for any person employed by or associated with any enterprise engaged in, or the activities of which affect, interstate or foreign commerce, to conduct or participate, directly or indirectly, in the conduct of such enterprise's affairs through a pattern of racketeering activity or collection of an unlawful debt.

19. 18 U.S.C. 1962(d) (1982).

(d) It shall be unlawful for any person to conspire to violate any of the provisions of subsections (a), (b), or (c) of this section.

20. 18 U.S.C. 1961(1) (1982) provides:

(1) "racketeering activity" means (A) any act or threat involving murder, kidnapping, gambling, arson, robbery, bribery, extortion, or dealing in narcotic or other dangerous drugs, which is chargeable under State law and punishable by imprisonment for more than one year; (B) any act which is indictable under any of the following provisions of title 18, United States Code: Section 201 (relating to bribery), section 224 (relating to sports bribery), section 471, 472, and 473 (relating to counterfeiting), section 659 (relating to theft from interstate shipment) if the act indictable under section 659 is felonious, section 664 (relating to embezzlement from pension and welfare funds), sections 891-894 (relating to extortionate credit transactions), section 1084 (relating to the transmission of gambling information), section 1341 (relating to mail fraud), section 1343 (relating to wire fraud), section 1510 (relating to obstruction of criminal investigations), section 1511 (relating to obstruction of State local law enforcement), section 1951 (relating to interference with commerce, robbery, or extortion), section 1952 (relating to racketeering), section 1953 (relating to interstate transportation of wagering paraphenalia), section 1954 (relating to unlawful welfare fund payments), section 1955 (relating to the prohibition of illegal gambling businesses), sections 2314 and 2315 (relating to interstate transportation of stolen property), sections 2341-2346 (relating to trafficking in contraband cigarettes), sections 2421-24 (relating to white slave traffic), (C) any act which is indictable under title 29, United States Code, section 186 (dealing with restrictions on payments and loans to labor organizations) or section 501(c) (relating to embezzelment from union funds), or (D) any offense involving fraud connected with a case under title 11, fraud in the sale of securities, or the felonious manufacture, importation, receiving, concealment, buying, selling, or otherwise dealing in narcotic or other dangerous drugs, punishable under any law of the United States.

21. 18 U.S.C. 1961, 1962 (1982).

"enterprise."[22]

The criminal penalties include imprisonment and fines as well as injunctive relief and divestiture.[23] The civil remedies include injunctive relief and divestiture.[24] The Attorney General of the United States may initiate a civil RICO action,[25] as can any person injured in his property or business.[26] "Person" includes individuals as well as entities capable of holding an interest in property.[27] Individuals bringing a civil action are entitled to recover treble damages should they prevail.[28]

B. The Call For Reform

There is a large and growing sentiment that the civil aspect of RICO is getting out of control in the number and variety of suits that have been brought under the statute.[29] It is argued that the current use of civil RICO was never intended nor envisioned to encompass such a wide range of civil actions.[30] The greatly expanded use of civil RICO is

---

22. 18 U.S.C. 1961, 1962 (1982).

23. 18 U.S.C. 1963 (1982).

24. 18 U.S.C. 1964(a) (1982).

(a) The district courts of the United States shall have jurisdiction to prevent and restrain violations of section 1962 of this chapter by issuing appropriate orders, including, but not limited to: ordering any person to divest himself of any interest, direct or indirect, in any enterprise; imposing reasonable restrictions on the future activities or investments of any person, including, but not limited to prohibiting any person from engaging in the same type of endeavor as the enterprise engaged in, the activities of which affect interstate or foreign commerce; or ordering dissolution or reorganization of any enterprise, making due provision for the rights of innocent persons.

25. 18 U.S.C. 1964(b) (1982).

(b) The Attorney General may institute proceedings under this section. In any action brought by the United States under this section, the court shall proceed as soon as practicable to the hearing and determination thereof. Pending final determination thereof, the court may at any time enter such restraining orders or prohibitions, or take such actions, including the acceptance of satisfactory performance bonds, as it shall deem proper.

26. 18 U.S.C. 1964(c) (1982).

(c) Any person injured in his business or property by reason of a violation of section 1962 of this chapter may sue therefor in any appropriate United States district court and recover threefold the damages he sustains and the cost of the suit, including a reasonable attorney's fee.

27. 18 U.S.C. 1961(3) (1982).

28. 18 U.S.C. 1964(c) (1982).

29. *See* Broucher, *Closing the Floodgates in the Aftermath of Sedima*, 31 N.Y.L. Sch. L. Rev. 133 (1986); LaCovara and Aronow, *The Legal Shakedown of Legitimate Business People: The Runaway Provisions of Private Civil RICO*, 21 New Eng. L. Rev. 1 (1985-86); Note, *Congress Responds to Sedima: Is There A Contract Out on Civil RICO?*, 19 Loy. L.A.L. Rev. 851 (1986).

30. *Id.*

probably due to its treble damages provision which is not only a great temptation in itself, but also carries great leverage for settlement value.[31] While sitting on the Ninth Circuit, Justice Kennedy wrote a concurring opinion in the case of *Schreiber Distribution Co. v. Serv-Well Furniture Co.*[32] in which he pointed to the great potential for misuse of civil RICO, saying:

> The potential range of criminal prosecution under the federal mail and wire fraud laws is vast, made so in part by expansive judicial interpretation. The reach of those statutes exists against a backdrop of prosecutorial discretion, however, discretion which, if sensitively exercised, operates as a check to the improvident exertion of federal power. No such check operates in the civil realm. A company eager to weaken an offending competitor obeys no constraints when it strikes with the sword of the Racketeer Influenced and Corrupt Organizations Act.
>
> It is most unlikely that Congress envisaged use of the RICO statute in a case such as the one before us, but we are required to follow where the words of the statute lead . . . . Thus, federal power inches forward when a statute is left unattended, whether from Congress' indifference or its acquiescence.[33]

Members of Congress have also recognized that RICO is exceeding its originally conceived limits and that there is a need to reform the statute. In the 99th Congress, three bills were introduced to reform RICO by making the elements more difficult to prove.[34] The bills were an attempt to make the RICO language clearer[35] and to reduce the stigma of the label of racketeer.[36]

In the 100th Congress, first session, a bill was introduced in the Senate which would limit recovery of treble damages by private individuals.[37] To recover treble damages, the individual would have to prove a prior criminal conviction of the RICO offenses.[38] Without a prior conviction, the individual would be limited to actual damages with punitive damages up to twice the amount of the actual damages if it

---

31. Note, *Congress Responds to Sedima: Is There A Contract Out On Civil RICO?*, 19 Loy. L.A.L. Rev. 851, 867-870 (1986).

32. 806 F.2d 1393 (9th Cir. 1986).

33. *Id.* at 1402 (Kennedy, Circuit Judge, concurring) (citations omitted).

34. S. 1521, 99th Cong., 1st Sess. (1985); H.R. 2517, 99th Cong., 1st Sess. (1985); H.R. 2943, 99th Cong., 1st Sess. (1985).

35. S. 1523, 100th Cong., 1st Sess. (1985).

36. H.R. 2517, 99th Cong., 1st Sess. (1985).

37. S. 1523, 100th Cong., 1st Sess. (1987).

38. *Id.* at C(2)(A).

is proven by clear and convincing evidence that "the defendant acted in conscious and wanton disregard of the consequences of the defendant's actions to the plaintiff."[39] The punitive damages are not allowed if federal or state securities laws already provide an available remedy, express or implied.[40] Additionally, punitive damages would only be awarded at the discretion of the trier of fact after taking eight enumerated balancing factors into account.[41] A similar bill was introduced in the House of Representatives with almost identical provisions as to treble damages.[42] Another bill introduced in the House kept the treble damages intact, but invoked a penalty for frivolous claims equal to treble the actual cost of the action, including reasonable attorney's fees.[43]

From recent law review articles and proposed legislation, it appears there is a movement afoot to significantly alter RICO to make it more difficult to bring an action and more difficult to recover treble damages. One method of accomplishing these purposes would be to increase the standard of proof to require clear and convincing evidence, a stricter standard than the ordinary civil standard of preponderance of the evidence for the predicate criminal acts.

The RICO statute does not specify the proper standard of proof to be applied to the predicate criminal acts. The correct standard of proof in civil RICO actions is properly a question for the courts to decide. As the Supreme Court stated in *Woodby v. INS*,[44] "[t]he question of what degree of proof is required is the kind of question which has been traditionally left to the judiciary to resolve."[45] The degree of proof that should be adopted is "clear and convincing evidence."

II. History of the Courts Addressing the Standard of Proof in Civil RICO

In 1985, the Supreme Court discussed the standard of proof for civil RICO actions in *Sedima, S.P.R.L. v. Imrex Co., Inc.*[46] Prior to *Sedima*, there was a split between the circuits over whether a prior criminal conviction for the predicate offenses was necessary before a civil action for treble damages could be initiated.[47] In deciding that a

---

39. *Id.* at C(2)(B)(ii)(II)(cc).
40. *Id.* at C(2)(b)(ii)(bb).
41. *Id.* at 3(A-H).
42. H.R. 2983, 100th Cong., 1st Sess. (1987).
43. H.R. 3240, 100th Cong., 1st Sess. (1987).
44. 385 U.S. 276 (1966).
45. *Id.* at 284.
46. 473 U.S. 482 (1985).
47. The split was between the Seventh and Second Circuit Courts of Appeal. Haroco, Inc. v. American Nat'l Bank & Trust of Chicago, 747 F.2d 384 (7th Cir. 1984),

prior conviction was not necessary to a private civil RICO action, the Court said:

> We are not at all convinced that the predicate acts must be established beyond a reasonable doubt in a proceeding under 1964(c). In a number of settings, conduct that can be punished as criminal only upon proof beyond a reasonable doubt will support civil sanctions under a preponderance standard. There is no indication that Congress sought to depart from this general principle here . . . . *But we do not decide the standard of proof issue today.*[48]

The proper standard of proof in a civil RICO action was not at issue in *Sedima* and the court properly labeled its statements as dicta.[49]

Following the decision in *Sedima*, several of the circuit courts addressed the standard of proof issue in dicta. In *United States v. Local 560 of International Brotherhood of Teamsters*,[50] the Third Circuit affirmed in dicta the district court's ruling that the standard of proof in civil RICO actions was by a preponderance of the evidence. In a footnote, the court gave the issue some analysis, basing its decision in part on the three part balancing test stated in *Santosky v. Kramer*[51] to determine whether a higher standard should be used.[52] The court referred to other situations in which higher standards of proof have and have not been used,[53] and factually compared them to the case at issue. Finally, the

---

held that it was not necessary to plead or prove a prior criminal conviction to recover treble damages in a civil RICO action. Sedima, S.P.R.L. v. Imrex Co., Inc., 741 F.2d 482 (2d Cir. 1984), held that a private action could only proceed against a defendant who had already been criminally convicted of the predicate acts.

48. *Sedima*, 473 U.S. at 491, (emphasis added) (citations omitted).

49. *Id.*

50. 780 F.2d 267 (3d Cir. 1985).

51. 455 U.S. 745, 754 (1982). The three part balancing test consisted of: "(1) the private interests affected by the proceeding; (2) the risk of error created by the state's chosen proceding; and (3) the countervailng government interest supporting use of the challenged procedure." The court in *Local 560* ruled that the defendants faced neither criminal sanctions nor any significant deprivations of liberty and that the nature of the remedy was remedial and not punitive, thereby reducing the risk of error, 780 F.2d 267 at 279-80 n.12, citing United States v. Cappetto, 502 F.2d 1351, 1358 (7th Cir. 1974), *cert. denied*, 420 U.S. 925 (1975); Farmers Bank of Delaware v. Bell Mortgage Corp., 452 F. Supp. 1278, 1280 (D. Del. 1978); Heinhold Commodities, Inc. v. McCarty, 513 F. Supp. 311, 313 (N.D. Ill. 1979).

52. *Local 560 of Int'l Bhd. of Teamsters*, 780 F.2d at 279 n.12.

53. Cited were: Santosky v. Kramer, 455 U.S. 745 (1982)(preponderance standard sufficient to balance defendant's risks in proceeding to terminate parental rights); Addington v. Texas, 441 U.S. 418 (1979)(Texas standard of clear, unequivocal and convincing in involuntary commitment proceedings is not violative of due process); Woodby v. INS, 385 U.S. 276 (1966)(clear, unequivocal and convincing standard of proof is necessary for deportation proceedings).

court cited the Supreme Court's dicta in *Sedima* as support for its decision.[54]

The Fifth Circuit addressed the standard of proof issue in *Armco Industrial Credit Corp. v. SLT Warehouse Co.*[55] The court adopted the preponderance standard in dicta relying on the "strong suggestion" of the Supreme Court in *Sedima*.[56] The Court in *Armco*, however, left the issue open by stating, "[s]ince the [Supreme] Court did not view the . . . standard of proof issue as 'close to the constitutional edge,' we are reluctant to inject another judicial construct into a statute that we are commanded to construe broadly."[57]

Both the Second[58] and Ninth[59] Circuits have stated in dicta that the preponderance standard is proper. The Ninth Circuit rested its decision on the dicta in *Sedima* and stated: "While there may be sound arguments that a greater burden of proof should apply, . . . we do not depart from the Supreme Court's directive in *Sedima* and the many courts which have applied the preponderance standard."[60] Additionally, several district courts have stated in dicta that a preponderance standard is appropriate.[61]

The first circuit court to rule on the standard of proof issue as a direct holding was the Seventh Circuit in *Liquid Air Corp. v. Rogers*.[62] The standard of proof was directly at issue and the court gave a fairly substantial analysis of its reasoning for holding for the preponderance standard.[63] The court cited the opinion in *Sedima* as well as the circuit and district courts which had concluded that preponderance was sufficient.[64] The court distinguished the cases which had held for a higher standard of proof when significant interests greater than money were

---

54. *Local 560 of Int'l Bhd. of Teamsters*, 780 F.2d at 280.

55. 782 F.2d 475 (5th Cir. 1986).

56. *Id.* at 481.

57. *Id.*

58. Cullen v. Margiotta, 811 F.2d 698, 731 (2d Cir.), *cert. denied*, 107 S. Ct. 3266 (1987).

59. Wilcox v. First Interstate Bank of Oregon, 815 F.2d 522, 531 (9th Cir. 1987).

60. *Id.* at 532.

61. Ford Motor Co. v. B & H Supply, Inc., 646 F. Supp. 975, 1001 (D. Minn. 1986); Bosteve, Ltd. v. Marauszwski, 642 F. Supp. 197, 202 n.7 (E.D.N.Y. 1986); Owl Constr. Co., Inc. v. Ronald Adams Contractor, Inc., 642 F. Supp. 475, 477 (E.D. La. 1986); Platsis v. E.F. Hutton & Co., 642 F. Supp. 1277, 1309 (E.D. Pa. 1986) *aff'd*, 829 F.2d 13, *cert. denied*, 108 U.S. 1227; Stainton v. Tarantino, 637 F. Supp. 1051, 1070 (E.D. Pa. 1986); Kimmel v. Peterson, 565 F. Supp. 476, 490 (E.D. Pa. 1983). In each of the above cited district court cases, the courts stated, in *dicta* only, that a preponderance standard was appropriate.

62. 834 F.2d 1297 (7th Cir. 1987).

63. *Id.* at 1302-1303.

64. *Id.* at 1302.

at stake.[65] Additionally, the court distinguished the common law fraud requirement that fraud needed to be proven by clear and convincing evidence because of the stigma attached to a fraud allegation.[66] The court held that the fraud issue was foreclosed by the Supreme Court's decision in *Herman & MacLean v. Huddleston*,[67] in which the Court held that proof by a preponderance of the evidence was sufficient in a securities fraud action.[68] The court also found that other federal statutes required proof by a mere preponderance.[69]

The Supreme Court has left the standard of proof issue open, at least as far as what the proper standard is below beyond a reasonable doubt. Five of the eleven circuit courts have indicated that the standard of proof in civil RICO is by a preponderance of the evidence. However, in four of the five circuit courts the issue was addressed in dicta.[70] The courts have given two basic reasons for determining that the proper standard of proof is by a preponderance. First, the RICO legislation is analogous to other federal legislation requiring only a preponderance. Second, the interests of a defendant in a civil RICO action are not so great as to require a higher standard of proof.

The standard of proof in civil RICO actions should be higher than the ordinary civil standard of a preponderance of the evidence. The interests of the defendant are higher than the interests of the plaintiff and are great enough to require that the plaintiff be burdened with a higher standard of proof. Additionally, there are significant differences in both the intent and effect of other federal legislation that the courts have cited as being analogous to the RICO legislation. These differences are great enough to distinguish the RICO standard of proof from the preponderance standard required by the other legislation. The proper standard of proof for the predicate criminal acts in a civil RICO action for treble damages should be by clear and convincing evidence.

III. Standards of Proof

A. *In General*

There are three basic standards of proof: (1) Beyond a Reasonable Doubt, used mainly in criminal cases; (2) By a Preponderance of the

65. *Id.* at 1302-1303.
66. *Id.* at 1303.
67. 459 U.S. 375 (1983).
68. *Liquid Air*, 834 F.2d at 1303 (citing Herman & MacLean v. Huddleston, 459 U.S. 375 (1983)).
69. *Id.* at 1303.
70. *See* note 11, supra.

Evidence, used mainly in civil cases for money damages; and (3) By Clear and Convincing Evidence, a standard used in civil cases where there is more at stake than money damages.[71]

In every case there is a risk that the factfinder may reach its verdict erroneously. The purpose of establishing a standard of proof is to balance the risk to the defendant of an erroneous verdict with a corresponding burden on the plaintiff.[72]

The first aspect of the risk of an erroneous verdict is the actual interests risked. In a civil action for compensatory damages, the interests risked by the plaintiff and the defendant are the same: the plaintiff risks the value of his injuries which is roughly equivalent to the amount of the damages. This is also what the defendant risks.[73] Thus, in this situation, it is only necessary that one party prevail by a preponderance since neither party runs a greater risk.

At the other extreme are criminal cases which require the highest standard of proof, beyond a reasonable doubt, because of the great risk that the defendant bears should the verdict be erroneous. Society has determined that the importance of ensuring that innocent people are not punished is much greater than the importance of letting a guilty person go unpunished. Hence, the defendant incurs a much greater risk of an erroneous verdict, which is balanced by a much higher standard of proof.[74]

Besides a comparison of the interests risked by the plaintiff and the defendant, another aspect to the risk placed upon the defendant is the likelihood that the type of allegation made will be true:

> In the absence of precedent, the likelihood that the *type of allegation* made by the party with the burden of persuasion is true controls how much proof must be adduced to meet that burden. Put more precisely, there is an inverse relationship between the likelihood that the *type of claim*, as opposed to the particular claim, which the burden-bearer is alleging is true and the degree of persuasiveness which will be demanded. Thus, if it is unlikely that a type of allegation can be supported, clear and convincing evidence will be required to meet the burden of persuasion.[75]

---

71. *See In re* Winship, 397 U.S. 358 (1970); United States v. Fatico, 458 F. Supp. 388 (E.D.N.Y. 1978).

72. *In re* Winship, 397 U.S. 358, 369 (1969) (Harlan, J., concurring).

73. *Id.*

74. *Id.* at 372.

75. General Motors Corp. v. Toyota Motor Co., Ltd., 467 F. Supp. 1142, 1173 (S.D. Ohio 1979).

Thus, the type of allegation that is made and the likelihood that the allegation will be true are factored into determining the burden of proof.

B. Clear and Convincing vs. Clear, Unequivocal and Convincing

Between the two extremes of the preponderance standard and the beyond a reasonable doubt standard there is a continuum of increasing probabilities of risk for the defendant.[76] Accordingly, there should be an increased burden on the plaintiff to compensate for the defendant's increased risk. The standard of proof used in this middle ground is some form of "clear and convincing" evidence.[77] Although the names differ between jurisdictions, there are basically two standards of proof between preponderance and beyond a reasonable doubt; "clear and convincing" and "clear, *unequivocal* and convincing."
The clear and convincing standard is used where moral turpitude is implied.[78] Examples of allegations involving moral turpitude include libel, fraud, securities fraud, and undue influence.[79] One federal court has quantified the certainty necessary for a factfinder to be convinced under the clear and convincing standard.[80] The factfinder must be certain to at least a 70% probability in order to find for the burdened party under a clear and convincing standard.[81] This compares to the greater than 50% certainty necessary to find under a preponderance standard. Besides quantification,

> [c]lear and convincing evidence is "that measure or degree of proof which will produce in the mind of the trier of facts a firm belief or conviction as to the allegations sought to be established. It . . . is more than a mere preponderance . . . [and] does not mean clear and unequivocal."[82]

Clear and convincing is to be distinguished from clear, *unequivocal* and convincing. "In situations where the various interests of society are pitted against restrictions on the liberty of the individual, a more demanding standard is frequently imposed, such as proof by clear, *unequivocal*, and convincing evidence."[83] A quantifiable amount that has

---

76. United States v. Fatico, 458 F. Supp. 388, 403 (E.D.N.Y. 1978).
77. *Id.* at 404-05.
78. *Id.*
79. *Id.* at 404.
80. General Motors Corp. v. Toyota Motor Co., Ltd., 467 F. Supp. 1142 (S.D. Ohio 1979).
81. United States v. Fatico, 458 F. Supp. 388, 405 (E.D.N.Y. 1978).
82. Hobson v. Eaton, 399 F.2d 781, 784 n.2 (6th Cir. 1968) (quoting Cross v. Ledford, 161 Ohio St. 469, 120 N.E.2d 118 (1954)).
83. *In re* Ballay, 482 F.2d 648, 662 (D.C. Cir. 1973)(emphasis added).

been used for the clear, unequivocal, and convincing standard is that the factfinder should be convinced of the truth of the proposition by more than an 80% probability.[84]

The Supreme Court has recognized the application of the clear, unequivocal, and convincing evidence standard to non-criminal cases where an individual's liberty interest is at stake. Examples of this include loss of citizenship and deportation proceedings.[85] The Supreme Court has also recognized the clear, unequivocal, and convincing standard as not violative of due process in state civil commitment cases.[86]

It is recognized that there are two middle standards of proof, the lower of which is clear and convincing. In the only two opinions which have attempted any analysis of the standard of proof for the predicate criminal acts in civil RICO (*United States v. Local 560*[87] and *Liquid Air Corp. v. Rogers*[88]), both courts have grouped the two standards under the one heading of "clear and convincing."[89] Both courts cited the Supreme Court cases of *Addington v. Texas*[90] and *Woodby v. INS*[91] as cases requiring only clear and convincing evidence for deprivation of important individual rights.[92] The cases cited required clear, unequivocal, and convincing evidence to deprive an individual of a liberty interest in a civil case.[93] These circuit courts, however, failed to recognize the difference between the two standards. Without recognizing the difference, it is impossible to correctly balance the interests risked with the proper standard of proof.

C. *Application to Civil RICO*

A civil RICO defendant incurs the risk of an erroneous verdict both in the likelihood that the claim asserted is true and in the interests risked. The elements of a civil RICO action are difficult at best to prove. Not only does the plaintiff have to show at least two seperate

---

84. *Fatico*, 458 F. Supp. at 405 (1978).
85. Woodby v. INS, 385 U.S. 276 (1966).
86. Addington v. Texas, 441 U.S. 418 (1978).
87. 780 F.2d 267 (3rd Cir. 1985).
88. 834 F.2d 1297 (7th Cir. 1987).
89. *Local 560 of Int'l. Bhd. of Teamsters*, 780 F.2d at 279-80 n.12; *Liquid Air Corp.*, 834 F.2d at 1302-03.
90. 441 U.S. 418 (1978).
91. 385 U.S. 276 (1966).
92. *Local 560 of Int'l Bhd. of Teamsters*, 780 F.2d at 279-80 n.12; *Liquid Air*, 834 F.2d at 1303.
93. In *Addington v. Texas*, the Court upheld a Texas standard of clear, unequivocal and convincing evidence for a civil committment proceeding. 441 U.S. at 433. In *Woodby v. INS*, the Court held that a clear, unequivocal and convincing standard was necessary to deport a legal alien. 385 U.S. at 286.

predicate acts within ten years of each other, but he also has to show that the two acts established a "pattern" and that they were committed within an enterprise separate and distinct from the defendant.[94] Establishing these elements in addition to establishing the predicate acts is more difficult than establishing the predicate acts alone. Therefore, it is less likely that the plaintiff will be able to meet his burden of persuasion. Accordingly, because of the type of action brought, the standard of proof for a civil RICO action should be higher than the standard of proof for the predicate acts alone.

In a civil RICO action for treble damages, there is no risk that the defendant will be deprived of his liberty: at most he will only lose money. Therefore, the clear, unequivocal, and convincing standard should not apply. However, if the plaintiff should prevail, he would be entitled to treble damages plus costs which includes reasonable attorney's fees.[95] Using a simple balancing test, the defendant risks more than three times the interest of the plaintiff in case of an erroneous verdict. This imbalance of risk should be compensated by a corresponding increase in the burden placed on the plaintiff. The clear and convincing standard would strike the proper balance as the defendant is running three times the risk of the plaintiff yet does not stand to be deprived of a liberty interest.

IV. Clear and Convincing Evidence for Civil RICO Because It Is An Allegation Involving Moral Turpitude

The increased risk to the defendant is not limited to money. The defendant risks the stigma of being labeled as a racketeer, a criminal, and/or one who commits fraud. These labels involve the moral turpitude of the defendant and are valid reasons for a corresponding increase to the plaintiff's burden by an increase in the standard of proof.[96]

A. Label of Racketeer

In *Exeter Towers Associates v. Bowditch*,[97] the court held that "Congress manifested an intent that these [civil RICO] remedies be available only in relation to activities having some association with 'racketeering' as that term is used in ordinary discourse."[98] The dictionary defines racketeer as one who operates a "racket" which is defined as: "3. An organized illegal activity, such as bootlegging or the extortion

94. 18 U.S.C. 1962 (1982).
95. 18 U.S.C. 1964(c) (1982).
96. United States v. Fatico, 458 F. Supp. 388, 404-05 (E.D.N.Y. 1978).
97. 604 F. Supp. 1547 (D. Mass. 1985).
98. *Id.* at 1553.

of money from legitimate businessmen by threat or violence. 4. Informal, a dishonest scheme, trick, business activity, etc.''[99]

It necessarily follows that one who has a civil RICO allegation brought against him stands a good chance of being labeled as one who participates in organized crime or dishonest schemes or tricks. The court in *United States v. Guiliano*[100] reversed a jury's guilty verdict on a bankruptcy fraud count because of the possible prejudice to the defendant resulting from a RICO count brought in conjunction with the bankruptcy fraud allegation.[101] The court stated:

> One of the hazards of a RICO count is that when the Government is unable to sustain a conviction under this statute, it will have to face the claim that the prejudicial effect of targeting a defendant with the label of 'racketeer' tainted the conviction on an otherwise valid count.[102]

Courts have recognized that there is a stigma attatched to the label of racketeer. In *Haroco v. American National Bank and Trust Co.*,[103] the Seventh Circuit held that in pleading a civil RICO action, the allegations in the complaint must pleaded with sufficient particularity so as to satisfy Federal Rules of Civil Procedure rule 9(b).[104] The reason for this requirement is to ensure that the defendant will not be stigmatized as a racketeer and one who commits fraud based on frivolous allegations in a complaint.[105] Addressing the strict pleading requirement under FRCP 9(b), the Court in *Exeter Towers* ruled that ''[t]he need for particularity of pleading is increased because of the 'in terrorem' effect that is associated with charges of racketeering activity.''[106] While particularity of pleading is not an issue in this Note, the requirement for particularity is an indication of the stigma attached to the label of racketeer.

The Supreme Court recognized that civil RICO actions stigmatize the defendant when it said in *Sedima*: ''As for stigma, a civil RICO action leaves no greater stain than do a number of other civil pro-

---

99. The Random House Dictionary of the English Language, Copyright 1973 by Random House, Inc., New York, NY.

100. 644 F.2d 85 (2d Cir. 1981).

101. *Id.*

102. *Id.* at 89.

103. 747 F.2d 384 (7th Cir. 1984), *aff'd*, 473 U.S. 606 (1985).

104. Fed. R. Civ. P. 9(b) states:

(b) Fraud, Mistake, Condition of the Mind. In all averments of fraud or mistake, the circumstances constituting fraud or mistake shall be stated with particularity. Malice, intent, knowledge, and other condition of mind of a person may be averred generally.

105. *Haroco*, 747 F.2d at 405.

106. Exeter Towers Assoc. v. Bowditch, 604 F. Supp. 1547, 1552 (D. Mass. 1985).

ceedings."[107] The Court did not explain the other civil proceedings to which it referred, but implicitly recognized that some "stain" is left on the defendant by a civil RICO action.

Two of the proposed pieces of legislation concerning RICO from the 100th Congress changed the phrase "racketeer influenced and corrupt organizations" to "pattern of unlawful activity."[108] This is evidence of the sensitivity that exists to labeling a party a "racketeer."

A civil RICO allegation suggests that the defendant is a "racketeer." The strong possibility of the stigma of this label is an unfair burden on the defendant. The risk of an unfair stigma on the defendant requires the balancing factor of a higher standard of proof, such as clear and convincing evidence, to be placed on the plaintiff.

B. Criminal Nature of RICO

1. Criminal Protection vs. A Higher Standard of Proof.—While on its face the RICO provisions for treble damages are civil, arguments have been made that the provisions are essentially criminal.[109] In order to bring a civil action, it is necessary to prove all of the elements for a criminal RICO action.[110] While a defendant to a civil RICO action cannot suffer criminal penalties, it is plain that it must be proven that he committed the criminal acts for the plaintiff to prevail.[111] The defendant must necessarily suffer the label of criminal without any of the safeguards of a criminal proceeding (including standard of proof).[112] It has been argued that one who does not incur the risk of criminal penalties is not entitled to criminal procedural protections.[113] However, that argument does not take into account the risk of the stigma of "criminal."

In *United States v. Cappetto*,[114] the Seventh Circuit denied criminal protections to a defendant in a civil RICO action stating, "[A] civil proceeding to enjoin those [predicate] acts is not rendered criminal in character by the fact that the acts are also punishable as crimes."[115]

---

107. Sedima, S.P.R.L. v. Imrex Co., Inc., 473 U.S. 479, 492 (1985).

108. S. 1523, 100th Cong., 1st Sess. (1987) and H.R. 2983, 100th Cong., 1st Sess. (1987).

109. Note, *Civil RICO Is A Misnomer: The Need For Criminal Procedural Protections In Actions Under 18 U.S.C. 1964*, 100 HARV. L. REV. 1288 (1987).

110. 18 U.S.C. 1964(c) (1982).

111. 18 U.S.C. 1964 (1982).

112. *See generally* Note, *supra* note 105, at 1291.

113. *See* United States v. Cappetto, 502 F.2d 1351 (7th Cir. 1974), *cert. denied*, 420 U.S. 925 (1975).

114. *Id.*

115. *Id.* at 1357.

The court in *United States v. Local 560*[116] cited *Cappetto* as precluding the "criminal nature" argument for a higher standard of proof in civil RICO.[117] However, the *Cappetto* case can be distinguished in several ways.

First, *Cappetto* was an injunctive proceeding brought by the United States government and not a private action for treble damages.[118] More importantly, the court was concerned with whether the defendant was entitled to *criminal procedural protections*.[119] The court only addressed the standard of proof issue insofar as to say that a defendant was not entitled to a criminal, "beyond a reasonable doubt," standard.[120] The argument here is not for criminal procedural protections but merely for a standard of proof that better reflects the risks that a defendant incurs in a civil RICO action, including the risk of being labeled as a criminal. The court in *Cappetto* seemed to recognize this distinction when it refused to declare the proper standard of proof, opting only to say that a defendant is not entitled to a criminal standard.[121]

While complete criminal protection may not be appropriate where no risk of criminal penalty exists, there needs to be some safeguard to protect against the incidental risks that go along with criminal accusations. Those risks should be balanced against the middle standard of proof: the clear and convincing standard.

2. The RICO Treble Damages Provision is Punitive.—In *Cappetto*, the court declared that civil RICO is remedial and not punitive.[122] However, the action in *Cappetto* was for injunctive relief and was taken by the United States government: it was not a private action for treble damages.[123]

Courts have been unsure whether punitive damages should be awarded in civil RICO actions for treble damages. The court, in *Moravian Development Corp. v. Dow Chemical Co.*,[124] held that punitive damages should not be awarded in civil RICO actions for treble damages.[125] The court compared civil RICO actions for treble damages to the treble damage remedy of the Sherman Antitrust Act. The Sherman Antitrust Act does not permit punitive damages in addition to the treble damages.[126]

---

116. 780 F.2d 267 (3d Cir. 1985).
117. *Id.* at 279, n.12.
118. *Cappetto*, 502 F.2d at 1354.
119. *Id.* at 1355 (emphasis added).
120. *Id.* at 1357.
121. *Id.*
122. *Id.*
123. *Id.* at 1359.
124. 651 F. Supp. 144 (E.D. Pa. 1986).
125. *Id.* at 150.
126. *Id.* at 149-150 (citing Arnott v. American Oil Co., 609 F.2d 873, 888 (8th Cir. 1979), *cert. denied*, 446 U.S. 918 (1980)).

The court, in disallowing punitive as well as treble damages in the civil RICO action stated: "The legislative history of RICO describes the RICO treble damages provision as 'another example of the antitrust remedy being adapted for use against organized criminality.'"[127]

Another district court has also been hesitant to award punitive damages on top of a civil RICO claim for treble damages. In *Al-Kazemi v. General Acceptance & Investment Corp.*,[128] the court denied a request for $150,000 in punitive damages stating, "[i]n view of the treble damages award for the RICO violation, an award of $50,000 in punitive damages will be sufficient to deter similar conduct by defendants in the future."[129]

The clear implication of the courts' decisions is that the treble damages provision of civil RICO is at least somewhat punitive. Hence, there is no need to award punitive damages in addition to the treble damages. That part of the treble damages which exceeds the value of the plaintiff's injury should rightly be considered punitive.

Some states already require a clear and convincing standard of proof for punitive damages levied on claims that require only a preponderance standard for compensatory damages.[130] Because the treble damages provision of civil RICO is punitive, the federal courts should take a cue from the state courts and require a clear proof for civil RICO actions brought for treble damages.

C. Fraud

In many jurisdictions fraud must be proven by a standard of proof higher than a mere preponderance of the evidence, usually clear and convincing evidence.[131] At common law, two forces were at work to

---

127. *Id.* at 150 (citing 116 CONG. REC. 35295 (1970)).

128. 633 F. Supp. 540 (D.D.C. 1986).

129. *Id.* at 544.

130. *District of Columbia*, Raynor v. Richardson-Merrell, Inc., 643 F. Supp. 238, 245 (D.D.C. 1986); *New York*, Roginsky v. Richardson-Merrell, Inc., 378 F.2d 832, 851 (2d Cir. 1967); *Indiana*, Travelers Indemnity Co. v. Armstrong, 442 N.E.2d 349, 363 (Ind. 1982); *Maine*, Tuttle v. Raymond, 494 A.2d 1353, 1363 (Me. 1985); *Wisconsin*, Chomicki v. Wittekind, 381 N.W.2d 561, 565, 128 Wis. 2d 188 (Wis. App. 1985).

131. At least thirteen jurisdictions require civil fraud to be proven by clear and convincing evidence. There is also federal authority which requires proving fraud by the clear and convincing standard. Lalone v. United States, 164 U.S. 255 (1896); United States v. American Bell Tel. Co., 167 U.S. 224 (1897); United States v. Hays, 35 F.2d 948 (10th Cir. 1929); Bowen v. B.F. Goodrich Co., 36 F.2d 306 (6th Cir. 1929); Northwestern Mut. Life Ins. Co. v. West, 62 D.C. App. 381, 68 F.2d 428 (D.C. Cir. 1933); Lopinto v. Haines, 185 Conn. 527, 441 A.2d 151 (1981); Weininger v. Metro. Fire Ins. Co., 359 Ill. 584, 195 N.E. 420 (1935); Harvey v. Phillips, 193 Ia. 231, 186 N.W. 910 (1922); Kansas Mill Owners & Mfr's Mut. Ins. Co. v. Rammelsberg, 58 Kan. 531, 50 Pac. 446 (1897); Gatchell v. Gatchell, 127 Me. 328, 143 A. 169 (1928); Conner v. Groh, 90 Md. 674, 45

bring about the higher standard of proof. First, there was a presumption that all men were honest.[132] To rebut that presumption a plaintiff had to show dishonesty by clear and convincing evidence.[133] Second, to label someone as having committed fraud was so onerous as to require a higher standard of proof to balance the increased risk.[134] The Federal Rules of Civil Procedure also recognize the danger to the defendant of the "fraud" label by requiring a higher pleading burden on the plaintiff for allegations of fraud.[135]

Several of the predicate criminal acts for a civil RICO action (and perhaps the most widely used) are based in fraud or *crimen falsi*.[136] These acts include counterfeiting, embezzlement, mail fraud, wire fraud, unlawful welfare fund payments, and securities fraud.[137]

Other federal statutes dealing with fraud use the clear and convincing standard of proof. The False Claims Act[138] allows the United States government to use fraud as a defense against parties bringing contract claims against it.[139] In order to use the defense, fraud must be proven by clear and convincing evidence.[140]

The dissent in *Wilcox v. First Interstate Bank*[141] (an action for treble damages predicated upon mail fraud) recognized three things about the majority ruling to support preponderance as the proper standard of proof

---

A. 1024 (1900); Wright v. Medlar, 176 Okla. 555, 56 P.2d 395 (1936); Metro. Casualty Ins. Co. v. N.B. Lesher Co., 152 Or. 161, 52 P.2d 1133 (1935); Pusic v. Salak, 261 Pa. 512, 104 A. 751 (1918); Bank of Pocahontas v. Ferimer, 161 Va. 37, 170 S.E. 591 (1933); Des Moines Auto Co. v. Tracy, 158 Wash. 23, 290 P. 423 (1930); Hunt v. Hunt, 91 W. Va. 685, 114 S.E. 283 (1922); *In re* Ball's Estate, Ball v. Boston, 153 Wis. 27, 141 N.W. 8 (1913).

132. *In re* Winship, 397 U.S. 358, 363 (1969).

133. *Id.*

134. *See generally id.* at 363-64.

135. *See supra* note 100.

136. *Crimen falsi* is defined as:

> Term generally refers to crimes in the nature of perjury or subornation of perjury, false statement, criminal fraud, embezzlement, false pretense, or any other offense which involves some element of deceitfulness, untruthfulness, or falsification bearing on witness' propensity to testify truthfully . . . . In the *civil law*, the crime of falsifying; which might be committed either by writing, as by the forgery of a will or other instrument; by words, as by bearing false witness, or perjury; and by acts, as by counterfeiting or adulterating the public money, dealing with false weights and measures, counterfeiting seals, and other fraudulent and deceitful practices.

BLACK'S LAW DICTIONARY 335 (5th ed. 1979).

137. 18 U.S.C. 1961(1) (Supp. 1987).

138. 31 U.S.C. 3729-33 (Supp. 1986).

139. Hageny v. United States, 570 F.2d 924 (Ct. Cl. 1978).

140. *Id.* at 933.

141. 815 F.2d 522 (9th Cir. 1987) (Boochever, J., dissenting).

in civil RICO cases: (1) the authority the majority relied upon was dicta; (2) the standard of proof for common law fraud was by clear and convincing evidence; and (3) Oregon followed the common law standard.[142] The dissent proposed that the standard of proof should be the same as for the predicate act standing alone where there was an express or implied private right of action.[143] But,

> [w]here there is no right to sue privately, courts should look to the most analogous common law action, usually a tort, for the appropriate burden of proof. Most states, but not all, require plaintiffs to establish fraud by clear and convincing evidence . . . . [I]f the burden of proof for RICO claims is by a preponderance of the evidence, [a defendant] is exposed to treble damages for the same acts that failed to establish liability at common law.[144]

Many civil RICO actions are predicated upon some type of fraud.[145] Because the clear and convincing standard of proof has been seen as essential to protecting a defendant's interests in civil fraud actions, a defendant's interests should likewise be protected in a civil RICO action predicated upon fraud. A civil RICO defendant is subjected to the same risk of the stigma of one who commits fraud. That risk is so onerous a burden that it demands a corresponding rise in the standard of proof—from the preponderance to the clear and convincing standard.

V. Civil RICO Is Distinguishable From Other Federal Legislation Which Offers Both Civil and Criminal Remedies; Specifically, Securities and Anti-Trust Legislation

The courts which have addressed the standard of proof issue in civil RICO have compared RICO to other federal legislation which requires only a preponderance standard. The court in *Liquid Air Corp. v. Rogers* ruled that clear and convincing evidence was not necessary in a civil RICO action based on mail and wire fraud because of the Supreme Court's decision in *Herman and MacLean v. Huddleston*,[146] a *securities fraud* action which held that preponderance was the proper standard of proof.[147] The court in *Liquid Air* acknowledged that treble damages is

142. *Id.* at 533.
143. *Id.*
144. *Id.* at 533-34.
145. *See* Arthur and White, *Civil RICO After Sedima: The New Weapon Against Business Fraud*, 23 Hous. L. Rev. 743, 767 (1986); Note, *Civil RICO and Its Application to "Garden Variety" Fraud within the Sixth Circuit*, 13 N. Ky. L. Rev. 463 (1987).
146. 459 U.S. 375 (1983).
147. *Liquid Air*, 834 F.2d 1297 at 1303 (citing *Herman & MacLean*, 459 U.S. 375).

not an available remedy under the securities laws and cited the treble damages remedy available under the Clayton Antitrust Act as additional authority for its position.[148] Treble damages actions under the antitrust laws require only a preponderance standard as per the Supreme Court's decision in *Ramsey v. United Mine Workers*.[149] The court in *United States v. Local 560* also cited *Herman & MacLean* in ruling for a preponderance standard of proof in civil RICO.[150] Federal securities legislation, as well as federal anti-trust legislation, is clearly distinguishable from RICO.

A. *Securities Legislation*

First, the securities legislation is distinguishable from RICO on its face in the remedies available. There is no treble damage provision for civil actions under the securities laws as there is in civil RICO.[151] Because much of the increased risk to the defendant comes from the treble damages remedy available in civil RICO, a statute without a treble damages remedy is clearly distinguishable as to the standard of proof.

Besides the facial differences of available remedies, the securities laws impose a higher standard of behavior on the securities industry. The Supreme Court held, in *Herman & MacLean v. Huddleston*,[152] that the proper standard of proof in a civil securities fraud action is by a preponderance of the evidence.[153] The Court distinguished the standard of proof in a civil securities fraud action from common law fraud. The Court found that one of the purposes of the Securities and Exchange Acts of 1933 and 1934 was to protect the public safety by imposing a higher standard of conduct on the securities industry.[154] The Court quoted itself from the case of *Blue Chip Stamps v. Manor Drug Stores*[155] when it said:

> [T]he typical fact situation in which the classic tort of misrepresentation and deceit evolved was light years away from the world of commercial transactions to which Rule 10b-5 is applicable. Moreover, the antifraud provisions of the securities laws

148. 834 F.2d 1297, 1303 (citing the Clayton Antitrust Act, 15 U.S.C. 12, 13, 14 to 19, 19a, 20, 21, 22 to 27).

149. 401 U.S. 302 (1970).

150. United States v. Local 560 of Int'l Bhd. of Teamsters, 780 F.2d 267, 279 n.12 (3rd Cir. 1985), *cert. denied*, 476 U.S. 1140 (1986).

151. *See generally* 15 U.S.C.

152. 459 U.S. 375 (1983).

153. *Id.* at 388 (1983).

154. *Id.* at 389.

155. 421 U.S. 723, 744-745 (1975).

> are not co-extensive with common law doctrines of fraud . . . . We therefore find reference to the common law in this instance unavailing.[156]

The court in *Huddleston* found that, given the congressional intent to impose a higher standard of conduct on the securities industry, the proper standard of proof was the preponderance standard.[157]

Previous to the *Huddleston* decision, the Supreme Court had been even stronger in ruling on the purpose of securities legislation passed after the stock market crash of 1929. In *Securites and Exchange Commission v. Capital Gains Research Bureau*,[158] the Court observed that "[a] fundamental purpose, common to these statutes [passed in the wake of the crash of 1929], was to substitute a philosophy of full disclosure for the philosophy of *caveat emptor* and thus to achieve a high standard of business ethics in the securities industry."[159]

The securities laws apply to a particular segment of society (the securities industry), and impose a much higher standard of conduct on that segment for particular reasons. RICO does not impose a higher standard of behavior on a given group of persons; the predicate acts are prohibited under criminal laws, and RICO is applicable to everyone. RICO is a system of remedies for criminal acts committed in a particular way.[160] Both from the actual provisions of RICO and the securities legislation as well as the ends to be achieved, the two pieces of legislation differ signficantly. Therefore, the standard of proof in a securities action is not analogous to the standard of proof in civil RICO.

B. Anti-Trust Laws

In holding for a preponderance standard of proof in civil RICO actions for treble damages, the Seventh Circuit in *Liquid Air Corp. v. Rogers* analogized the civil RICO remedies to the civil remedies of the antitrust laws which allow for treble damages and require only a preponderance standard of proof.[161] However, the antitrust legislation is distinguishable from civil RICO in its purpose and use.

The court in *Metro Cable Co. v. CATV of Rockford, Inc.*[162] held that the purpose of the antitrust laws was to regulate commerce so as to prevent monopolies.[163] Activities which fall outside of the area of

---

156. 459 U.S. at 388-9.
157. *Id.* at 389.
158. 375 U.S. 180 (1963).
159. *Id.* at 186.
160. *See generally* Title 18 United State Code.
161. 834 F.2d at 1303.
162. 516 F.2d 220 (7th Cir. 1975).
163. *Id.* at 227.

commerce and trade are not regulated by the antitrust laws "even though the purpose and effect of the concerted activities is to eliminate competition."[164] The Supreme Court's holding in *Ramsey v. United Mineworkers*, that a mere preponderance was the proper standard of proof for a treble damages action under the antitrust laws,[165] should be balanced against Congress' strong purpose of regulating commerce in order to prevent restraint of trade.

The RICO provisions for private civil actions were designed as another tool to use against organized crime.[166] However, the *de facto* use of the treble damages remedy by private civil litigants has been as a new remedy for ordinary business fraud, and not primarily as a weapon against organized crime.[167]

The analogy between the two statutes is that Congress' purpose in enacting each statute is strong enough to override the added risks on the defendant. Therefore, those additional risks do not need to be balanced against a higher standard of proof: a preponderance standard will suffice. However, with the original purpose of the private civil remedies available under RICO having *de facto* changed, there is no longer a compelling reason not to protect the defendant from the additional risks he incurs. Because the original purpose of civil RICO no longer exists to balance the preponderance standard, the analogy between RICO and the antitrust laws fails as to the standard of proof. The proper standard of proof in civil RICO actions for treble damages remains the clear and convincing standard.

VI. Conclusion

There is no doubt that organized crime is a problem that should be addressed by giving prosecutors the necessary tools to combat it. There is also good reason for making these tools broad and flexible and for putting remedies against such crime into the hands of private civil litigants. The RICO legislation goes a long way toward accomplishing these goals. However, the ends never justify the means where the means are unfair and unjust. Granting private civil litigants the power to bring treble damages actions within the broad scope of RICO places too great a risk on the defendant. These risks include the lower likelihood that a RICO action can be successfully brought, the increased interests risked

164. *Id.* at 227-228.

165. 401 U.S. 302 (1971).

166. *See generally* Lacovara and Aronow, *The Legal Shakedown of Legitimate Business People: The Runaway Provisions of Private Civil RICO*, 21 New Eng. L. Rev. 1 (1985-86).

167. *Id.* at 13-23.

by the defendant from the treble damages, and the risk to the defendant's reputation from allegations which involve moral turpitude. These risks must be balanced with an appropriate standard of proof, which in this case, is the standard of clear and convincing evidence.

Christopher M. Maine